DOCUMENT RESUME

ED 218 922 FL 012 807

AUTHOR Alderman, Donald L; Holland, Paul W.
TITLE Item Performance Across Native Language Groups on the Test of English as a Foreign Language. TOEFL Research Reports, 9.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-81-16
PUB DATE 81
NOTE 149p.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Comparative Testing; *English (Second Language); Factor Analysis; Foreign Students; *Language Proficiency; Postsecondary Education; *Reading Tests; Statistical Data; *Test Validity
IDENTIFIERS *Test of English as a Foreign Language

ABSTRACT

The Test of English as a Foreign Language (TOEFL) was examined for instances in which the item performance of examinees with comparable scores differed according to their native languages. A chi-square procedure, sensitive to deviations of less than ten percent from the expected frequencies of correct item responses across several language groups, revealed significant differences on seven-eighths of the TOEFL items. Reviewers familiar with particular languages could attribute the relative advantage or disadvantage of those language groups on a specific item to linguistic similarities or dissimilarities with the English language. Reviewers could not, however, identify which items would exhibit differential performance across groups based upon inspection of a test form and answer key alone. These findings suggest that examinees' performance on given items in a test of proficiency in a second language will vary according to linguistic contrasts with their native language and that statistical procedures will be necessary for identifying items with exaggerated or unexplained differences across language groups. Four appendices provide a large number of statistical tables and graphs. (Author/AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218922

Item Performance Across Native Language Groups on the Test of English as a Foreign Language

Donald L. Alderman

and

Paul W. Holland

Educational Testing Service
Princeton, N.J.

RR 81-16

TOEFL Research Reports, 9.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Charles Stansfield
ETS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

FL 012 807

The Test of English as a Foreign Language (TOEFL) was developed in 1963 by a National Council on the Testing of English as a Foreign Language, which was formed through the cooperative effort of over thirty organizations, public and private, that were concerned with testing the English proficiency of non-native speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program and in 1973 a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; Graduate Record Examinations Board members are associated with graduate education.

ETS administers the TOEFL program under the general direction of a Policy Council that was established by, and is affiliated with, the sponsoring organizations. Members of the Policy Council represent the College Board and the Graduate Record Examinations Board and such institutions and agencies as graduate schools of business, junior and community colleges, nonprofit educational exchange agencies, and agencies of the United States government.

Copyright © 1980 by Educational Testing Service. All rights reserved.

Unauthorized reproduction in whole or in part is prohibited.

Contents

	Page
Introduction	1
Method	4
Samples	5
Data Analysis	6
Specialist Reviews	10
Results and Discussion	11
Language Groups	11
Test Performance and Item Difficulties	12
Differential Item Performance Across Language Groups	18
Linguistic Explanations for Discrepant Item Performance	27
Summary and Conclusions	31
References	33
Appendices	37
A. Descriptive Profiles of Language Groups	37
B. Illustrative Delta Plots of Item Difficulty Values--TOEFL Section III: Reading Comprehension and Vocabulary, Administration I	43
C. Score Distributions by Test Section and Language Group	67
D. Percentage Deviations from Expected Frequencies of Correct Item Responses	75

List of Tables

Table		Page
1.	Means and Standard Deviations of Scores by TOEFL Sections Across Language Groups	14
2.	Means, Standard Deviations, and Correlations of Item Difficulties (Deltas) Across Language Groups: Administration I	15
3.	Means, Standard Deviations, and Correlations of Item Difficulties (Deltas) Across Language Groups: Administration II	16
4.	Identification of Discrepant Item Performance by Chi-Square Index	20
5.	Percentage Deviations from Expected Frequencies of Correct Item Responses Section III: Reading Comprehension and Vocabulary Administration I	22
6.	Distribution of Percentage Deviations from Expected Frequencies of Correct Item Responses by Language Groups: Administration I	25
7.	Distribution of Percentage Deviations from Expected Frequencies of Correct Item Responses by Language Groups: Administration II	26

INTRODUCTION

Whether some test items offer an unfair advantage or disadvantage for particular groups of examinees is a difficult question. It poses difficulties in detecting those items on which examinees' performance varies by group rather than by overall ability, in determining the extent of the discrepancy in item performance among groups, and in interpreting the reasons for discrepant item performance by groups. The present study addresses these problems within the context of the performance of foreign language groups on the Test of English as a Foreign Language (TOEFL), a test of proficiency in the English language which is taken annually by approximately 250,000 foreign applicants to colleges and universities in the United States. Because discrepant item performance on a test of English proficiency may be related to linguistic similarities and dissimilarities with foreign examinees' native languages, this context would seem to offer an unusual opportunity to tie discrepant item performance to specific features of the group's native language.

There have been several recent studies comparing statistical procedures for detecting item bias (Ironson & Subkoviak, 1979; Rudner, Getson & Knight, 1980a, 1980b; Shepard, Camilli & Averill, 1980). Such procedures include analysis of variance of item difficulty values by groups (e.g., Cleary & Hilton, 1968), plots of item difficulty values by groups (e.g., Angoff & Ford, 1973; Angoff & Sharon, 1974), comparisons of item characteristic curves by groups (e.g., Lord, 1980), and chi-square analysis of item response frequencies by groups (e.g., Scheuneman, 1979).

Each of these techniques has certain strengths and weaknesses in terms of its sensitivity to discrepant item performance, its susceptibility to variation due to irrelevant factors, and its ease of implementation as a routine part of the test development process. No single statistical procedure, however, seems to be fully satisfactory though each may be adequate for a given situation. The chi-square analysis, the statistical procedure chosen for this study, lends itself readily to a context involving several different groups and adapts easily to yield a measure of the extent of discrepant item performance among groups.

Despite the widespread interest in statistical procedures for detecting instances of possible item bias, there has been little attention devoted to developing meaningful indices of the extent to which groups of comparable ability differ with respect to their performance on specific test items. It has perhaps been implicit that the magnitude of statistical tests of significance or the probability level of such significance tests convey sufficient information about an item. And certain statistical procedures suggest their own indices: differences in the difficulty parameter (b) of groups' item characteristic curves (Lord, 1977, 1980) and perpendicular distances to the line of best fit in plots of item difficulty values (Angoff & Sharon, 1974; Sinnott, 1980), for example. Yet differential item performance expressed as a difference in b parameters or as a distance in delta units may fail to convey the extent of a group's relative advantage or disadvantage on a scale familiar to concerned audiences (e.g., test developers, policy makers, group members). Taking simple differences in the percentage of correct item

responses between groups is a misleading alternative since it assumes equal ability among groups regardless of each group's prior experience, instruction, or other possible dimensions of natural and cultural differences among groups which might affect ability. Stricker (1981) notes limitations of current statistical procedures and proposes the use of a partial correlation coefficient for reflecting the strength of association between item performance and group membership with total test score held constant. But the magnitude of this index would fluctuate according to the proportions of examinees passing and failing an item and would be insensitive to changes in an item's discrimination power at different levels of ability (Stricker, 1981). The present study involves several foreign language groups and simply extends chi-square analysis to provide a measure of discrepant item performance by comparing the observed frequency of correct item responses with the expected frequency of correct item responses by group. The percentage deviation of the observed from the expected frequency of correct item responses describes a group's relative advantage or disadvantage on a scale that seems inherently meaningful.

Even with an appropriate statistical procedure for detecting discrepant item performance and some measure for determining the extent of a group's relative advantage or disadvantage, it is still often difficult to identify the reasons for the group differences found. Scheuneman (in press) reminds us that early studies of item bias probably began with the assumption that the reasons for differential item performance would be self-evident once items had been earmarked by statistical procedures. But such revelations following closer scrutiny of item

content have been the exception rather than the rule. Angoff (in press) suggests that statistical procedures simply identify items on which there is differential perform nce by groups, not necessarily items on which there exist clear contextual cues or any other evident explanation for observed discrepancies in group performance. Whether statistical or contextual evidence is alone sufficient to demonstrate item bias seems a moot point. On a test of English proficiency there should certainly be discrepant item performance across foreign language groups that have different degrees of affinity to the English language. This study examined procedures for uncovering instances of discrepant item performance using both statistical and judgmental approaches.

The study thus had three primary objectives: (a) to detect instances of differential item performance across foreign language groups on the TOEFL through the use of an appropriate statistical procedure; (b) to express each group's relative advantage or disadvantage on TOEFL items on a meaningful scale; and (c) to examine the linguistic features of discrepant items for possible generalities applicable in subsequent test development. This follows and extends the earlier research of Angoff and Sharon (1974) which dealt with differential item performance on the TOEFL through analysis of variance and plots of item difficulty values.

METHOD

Samples of test candidates representing six different foreign language groups were drawn from two international administrations of the TOEFL. A chi-square analysis by language group, correct-incorrect item response, and score level was performed for each of the three distinct TOEFL sections for each test administration. Specialists in English as a foreign

language reviewed the results from the first test administration regarding discrepant item performance and language groups' percentage deviations from expected frequencies of correct item responses, and then attempted to identify probable instances of discrepant item performance on a second test form from a judgmental rather than statistical perspective based on their reviews and expertise.

Samples

The language groups included test candidates whose native languages were: African; Arabic; Chinese; Germanic; Japanese; and Spanish. The latter five native languages were also in the study of item-group interactions by Angoff and Sharon (1974) and all six languages were in the recent study of the test's factor structure by Swinton and Powers (1980). These particular language groups had been chosen both for their diversity and for their volume of test candidates. Arabic, Chinese, and Japanese represent language groups with high numbers of TOEFL candidates. Germanic and Spanish represent language groups with strong linguistic similarities to the English language. The African language group offers further opportunity for linguistic contrasts across groups.

Samples of approximately 1,000 test candidates from each language group were drawn from data files for two international TOEFL administrations given in November 1976 (Form YTF4) and November 1979 (Form BTF11). The analysis of two test forms provided a check on the stability of findings for discrepant item performance by forms. Similarly, for both administrations, a second sample of Chinese candidates was drawn in order to permit the checking of results for fluctuations which might arise from sampling alone. The African samples actually included four languages in the

Guinea-Senagalese region (i.e., Ibo, Yoruba, Efik, and Fanti) and the Germanic samples included a number of related native languages (i.e., German, Dutch, Swedish, and Danish) so that minimum sample size would be close to 500 with all test candidates from these groups.

Data Analysis

The chi-square analysis performed in this study was a modification of the procedure described by Scheuneman (1979). Scheuneman's statistic is a direct function of the proportion of correct item responses (Scheuneman, 1979, 1981) and does not have a null distribution that follows a chi-square statistic for large samples (e.g., Baker, 1981; Shepard et al., 1980). The frequency of both correct and incorrect item responses should enter into calculation of the appropriate chi-square statistic.

Detecting discrepant item performance necessarily involves an analysis of item responses by groups. Groups may differ in the relevant ability and examinees at different levels of ability should exhibit different frequencies of correct item responses. Clearly then statistical procedures for detecting discrepant item performance must take ability level into consideration. In our chi-square analysis this is accomplished by contrasting the observed and expected frequencies of item responses across groups within score strata. The distribution of scores is broken down into a reasonable number of strata, where a larger number of strata offer greater control for overall group differences in ability under the assumption that the scores indeed represent a valid measure of the same construct across groups.

The total chi-square value calculated for each item in this study took the form:

$$\hat{\chi}^2_{\alpha} = \sum_i \sum_j \sum_k \frac{\left(n_{ijk} - \frac{n_{ij+} \times n_{i+k}}{n_{i++}}\right)^2}{\left(\frac{n_{ij+} \times n_{i+k}}{n_{i++}}\right)}$$

where α refers to a specific item within each test section of the three separate TOEFL Sections on Listening Comprehension ($\alpha = 1,2,\ldots, 50$), Structure and Written Expression ($\alpha = 1,2,\ldots, 40$), and Reading Comprehension and Vocabulary ($\alpha = 1,2,\ldots,60$);

i refers to a given score stratum for a test section, with approximately equal percentages of test candidates within each stratum ($i = 1,2,\ldots,10$);

j refers to a particular group, here six different foreign language groups as well as a duplicate sample for one group ($j = 1,2,\ldots,7$);

k refers to a particular item response, here correct or incorrect ($k = 0$ incorrect, 1 correct);

n_{ijk} is the frequency of test candidates within the i^{th} score stratum from the j^{th} language group with response k;

n_{ij+} is the total frequency of test candidates within the i^{th} score stratum from the j^{th} language group, regardless of item response;

n_{i+k} is the total frequency of test candidates within the i^{th} score stratum with item response k, regardless of language groups; and

n_{i++} is the total frequency of test candidates within the i^{th} score stratum, regardless of language group or item response.

It should be clear that n_{ijk} is simply an observed cell frequency in a three-way classification table of item responses, language groups, and score strata. Similarly, n_{ij+}, n_{i+k}, and n_{i++} represent marginal totals of these cell frequencies in the same three-way classification table. The degrees of freedom for each χ^2_α would be equal to $S(L-1)(C-1)$, where S represents the total number of score strata, L the total number of groups, and C the total number of possible item responses. This approach also permits checking for discrepant item performance at a particular score stratum by omitting the summation over strata and examining each of the S_i values for χ^2_α (i) as a chi-square statistic with (L-1)(C-1) degrees of freedom.

While n_{ijk} denotes the observed frequency of item performance by language group and score stratum, the subtrahend of the numerator and the denominator in the equation for χ^2_α denote the expected frequency of item performance for the same language group and score stratum:

$$\hat{n}_{ijk} = \frac{n_{ij+} \times n_{i+k}}{n_{i++}}$$

$$= n_{ij+} \times \frac{n_{i+k}}{n_{i++}}$$

The first term, n_{ij+}, is the total number of test candidates from a particular language group, L_j, within a given score stratum, S_i; the second term, $\frac{n_{i+k}}{n_{i++}}$, is independent of language groups and represents the proportion of test candidates in the total sample from that same score stratum with item response k. The product of these two terms must then be

the expected frequency of item performance for a single cell in the three-way classification table.

Denoting correct item responses by the substitution of the subscript 1 for k, the observed and expected frequencies of correct item responses for the i^{th} score stratum, S_i, and j^{th} language group, L_j, can be written as n_{ij1} and $\hat{n}_{ij1}$. The value for n_{ij1} is an observed frequency and for $\hat{n}_{ij1}$ follows from the equation for $\hat{n}_{ijk}$. Deviations from expected frequencies of correct item responses for a given L_j become:

$$d_{j1} = n_{+j1} - \hat{n}_{+j1}$$

This is easily expressed as a percentage deviation from the expected frequency of correct item responses for a particular language group, L_j, by collapsing across score strata and dividing by expected frequencies, with multiplication by a constant for conversion into percentages:

$$b_{j1} = \frac{d_{j1}}{\hat{n}_{+j1}} \times 100$$

For correct item responses, b_{j1} reflects the relative advantage or disadvantage for a particular language group in comparison with the total sample as a simple percentage. There is such a percentage deviation from expected frequency of correct item responses for each language on each test item.

These analyses for discrepant item performance, both χ^2_α and b_{j1}, aggregate across score strata and thereby make an implicit assumption regarding the stability of relative advantages and disadvantages across strata. This assumption is equivalent to the assertion that there is no three-way interaction among item response, language group, and score

stratum (i.e., the relative advantage or disadvantage of a group on an item is the same in each score stratum). Likelihood-ratio chi-square values for interaction were computed according to the log-linear models described in Bishop, Fienberg and Holland (1975).

Probabilities of correct item responses within groups were also converted to the delta scale according to an inverse normal transformation (see Angoff & Sharon, 1974). The area of the normal curve, N(13,4), above a given point on this scale corresponds to the probability of a correct response. This transformation facilitates the interpretation of the correlations of item difficulty values between groups. The correlations present another description of the agreement in item difficulties among different groups.

Specialist Reviews

The results of the above analyses from the first test administration were submitted to specialists in English as a second or foreign language for review. These persons were university faculty members who served on committees responsible for research and test development activities for the TOEFL. Their task was to suggest plausible explanations for the discrepant item performance on the ten items within each TOEFL section that had the highest chi-square values. The information made available for the review included the test form and answer key, means and standard deviations of section scores by language groups, item difficulty values by language groups, chi-square values by items, the rank-order of items within sections according to their total chi-square values, and percentage deviations from expected frequencies of correct item responses by language groups (see Appendix D).

After examining these materials and suggesting plausible explanations for discrepant item performance, reviewers turned to a second test form and answer key and, on the basis of their experience with the first form and their expertise in the field, attempted to select the ten items from each section which would in their judgment show the largest relative advantages or disadvantages across the same language groups. The only information made available for this second review was the test form and answer key. Reviewers had to use item content and their knowledge of linguistic similarities and dissimilarities, as well as findings from the first form, in order to identify potential instances of discrepant item performance.

RESULTS AND DISCUSSION

Language Groups

The samples of test candidates whose native languages were Chinese, Japanese, and Spanish approached 1,000 examinees for each language group on both test administrations. The African, Arabic, and Germanic groups included all examinees with the appropriate native languages for the first test administration. For the second test administration it was possible to sample approximately 1,000 examinees whose native language was Arabic, but the African and Germanic groups again included all examinees from the designated native languages.

These groups differed in other respects besides their native languages. Naturally there were differences in native country both across and within language groups. Nearly all test candidates in the Japanese group came from Japan and in the African group from Nigeria. The native countries for examinees in the Germanic group (i.e., Germany, Netherlands, Sweden,

and Denmark), closely followed their respective native languages. The Arabic group came primarily from Jordan, Lebanon, and Egypt. The dominant native country for Chinese test candidates was Hong Kong in the first test administration and Taiwan in the second test administration.[1] Candidates' reasons for taking the TOEFL varied somewhat by administration as well as by language group; candidates usually cited application to either an undergraduate college or a graduate school as their reason for taking the test. Only the African language group showed a clear majority (85%) applying to undergraduate colleges in both administrations. Close to ninety percent of the African and Arabic test candidates were males; the other language groups had a less dramatic majority of male test candidates with the highest percentage of female test candidates in the Germanic sample (44%). There were even differences among the language groups in their prior exposure to the TOEFL, with approximately one-fourth of the Arabic, Chinese, and Japanese samples reporting a previous TOEFL administration. Appendix A presents descriptive profiles of the language groups with regard to these characteristics.

Test Performance and Item Difficulties

The test performance of the language groups is reported in Table 1 in terms of the number of correct item responses by test section. The order of the language groups in this table follows their rank-order in total test score. It is apparent that the two language groups with the closest affinity to the English language, Germanic and Spanish, attained the highest test scores. The high mean scores of the African group on Structure

---

[1]Taiwanese test candidates had been omitted from the first sample due to the conditions of test administrations during that time period in Taiwan.

and Written Expression and on Reading Comprehension and Vocabulary, however, is probably less a function of linguistic similarities with the English language than of extensive instruction in and through English, usually for three years in primary school and five years in secondary school. The low mean score of the African group on Listening Comprehension perhaps attests to their familiarity and use of different accents and speech patterns than those prevalent in English as spoken in the United States. That the rank-order of language groups varies by section scores lends support to the existence of separate dimensions in English proficiency and to the examination of discrepant item performance by section. The similarities of the means and standard deviations for the dual Chinese samples drawn for each test administration suggest little fluctuation due to sampling within test administrations. The standard deviations of the language groups' section scores reflect sometimes marked differences in the spread of the respective groups' score distributions (see Appendix C).

Correlations of the item difficulty values between groups appear in Table 2 for the first test administration[2] and Table 3 for the second test administration. The means of the item difficulty values reported in these tables restate the groups' relative standing on each test section with respect to their mean scores, as given in Table 1. The correlations of item difficulties for the two Chinese samples on each test section, $r = .99$, show the strength of association between independent samples

---

[2]One item from the TOEFL Section on Structure and Written Expression had been omitted from the operational scoring for this test form because of differences between American and British vocabularies. It was also excluded from analyses in this study.

Table 1

Means and Standard Deviations of Scores by TOEFL Sections Across Language Groups

Language Group	N	Section I Listening Comprehension Mean	sd	Section II Structure and Written Expression Mean	sd	Section III Reading Comprehension and Vocabulary Mean	sd
Administration I							
Germanic	596	43.61	5.5	30.65	5.2	47.95	7.1
Spanish	991	34.54	9.7	24.73	7.3	44.79	9.0
African	740	29.51	8.5	28.77	6.1	41.24	9.1
Chinese 1	998	31.50	8.8	24.79	6.4	37.25	9.9
Chinese 2	998	31.88	8.5	24.64	6.4	37.32	9.9
Japanese	997	30.56	8.6	22.79	6.3	36.31	10.1
Arabic	686	32.36	9.6	22.98	6.9	35.00	10.5
Administration II							
Germanic	960	42.91	5.1	34.04	4.6	45.80	7.2
Spanish	992	34.88	9.4	27.53	7.8	40.72	9.4
African	449	27.48	7.7	29.98	5.7	39.97	8.0
Chinese 1	993	29.49	7.8	26.31	5.5	38.15	8.5
Chinese 2	996	28.97	7.7	26.10	5.6	38.17	8.2
Japanese	996	30.14	8.1	26.46	5.9	33.73	9.5
Arabic	987	30.38	9.1	23.97	7.2	30.99	9.6

Table 2

Means, Standard Deviations, and Correlations of Item Difficulties (Deltas) Across Language Groups: Administration I

Language Group	Number of Items	Mean Delta	sd	Item Difficulty Correlations					
				Spanish	African	Chinese 1	Chinese 2	Japanese	Arabic
Section I: Listening Comprehension									
Germanic	50	7.93	2.0	.87	.73	.81	.80	.74	.85
Spanish	50	10.56	1.8		.75	.86	.87	.84	.90
African	50	11.99	1.9			.82	.82	.73	.73
Chinese 1	50	11.52	2.0				.99	.86	.88
Chinese 2	50	11.43	2.0					.86	.88
Japanese	50	11.70	2.1						.82
Arabic	50	11.35	1.8						
Section II: Structure and Written Expression									
Germanic	39	9.15	2.7	.87	.69	.81	.80	.79	.89
Spanish	39	11.47	2.1		.69	.74	.74	.80	.88
African	39	10.02	2.3			.63	.64	.59	.79
Chinese 1	39	11.39	2.3				.99+	.84	.88
Chinese 2	39	11.44	2.3					.84	.88
Japanese	39	12.04	2.4						.59
Arabic	39	11.93	2.3						
Section III: Reading Comprehension and Vocabulary									
Germanic	60	8.93	2.8	.67	.54	.80	.79	.77	.83
Spanish	60	9.90	2.4		.55	.44	.43	.65	.60
African	60	10.81	2.0			.57	.56	.52	.61
Chinese 1	60	11.63	2.1				.99+	.73	.83
Chinese 2	60	11.62	2.1					.73	.82
Japanese	60	11.78	2.2						.84
Arabic	60	12.04	2.1						

Table 3

Means, Standard Deviations, and Correlations of Item Difficulties (Deltas) Across Language Groups: Administration II

Language Group	Number of Items	Mean Delta	sd	Item Difficulty Correlations					
				Spanish	African	Chinese 1	Chinese 2	Japanese	Arabic
Section I: Listening Comprehension									
Germanic	50	8.16	2.1	.88	.71	.73	.70	.70	.82
Spanish	50	10.74	1.7		.68	.85	.83	.81	.92
African	50	12.44	1.8			.71	.68	.69	.69
Chinese 1	50	11.99	1.9				.99	.87	.79
Chinese 2	50	12.10	1.9					.81	.92
Japanese	50	11.80	2.0						.78
Arabic	50	11.82	1.7						
Section II: Structure and Written Expression									
Germanic	40	8.20	2.2	.87	.76	.70	.68	.78	.78
Spanish	40	10.83	1.9		.76	.75	.74	.81	.83
African	40	9.91	2.2			.69	.67	.75	.84
Chinese 1	40	11.10	2.7				.99+	.92	.89
Chinese 2	40	11.17	2.6					.92	.89
Japanese	40	11.00	2.7						.93
Arabic	40	11.85	2.2						
Section III: Reading Comprehension and Vocabulary									
Germanic	60	9.30	3.1	.76	.70	.83	.84	.81	.82
Spanish	60	10.72	2.7		.75	.81	.81	.78	.84
African	60	10.82	3.0			.79	.79	.75	.78
Chinese 1	60	11.35	2.5				.99+	.92	.91
Chinese 2	60	11.35	2.5					.92	.91
Japanese	60	12.24	2.4						.89
Arabic	60	12.77	2.3						

from the same language group. These duplicate samples also have nearly identical correlations with the item difficulties of other language groups. But there is certainly variation in the strength of association between item difficulty values across groups as reflected by these correlations with the Chinese samples, ranging from a low correlation between the Chinese and Spanish groups on Section III in Administration I, $r = .43$, to a high correlation between the Chinese and Japanese groups on Section III in Administration II, $r = .92$.

The lower correlations in Tables 2 and 3 indicate a greater relative success on some items and failure on other items for one language group contrasted with another. Groups with disparate mean scores can have similar patterns of item difficulty values, for example, the Germanic-Arabic correlation on Section II in Administration I, $r = .89$. And groups with comparable mean scores can have dissimilar patterns of item difficulty values, for example, the Japanese-Arabic correlation on Section II in Administration I, $r = .59$. The lower correlations reflect the degree of dissimilarity in patterns of item difficulties between two groups; high correlations, on the other hand, may reflect consistent differences in item difficulty values across items. Appendix B illustrates relative advantages and disadvantages in item performance through the delta plots corresponding to the correlations for Section III in Administration I, the section with the lowest correlations and the greatest range in correlation values.

While the average correlations of item difficulty values for the TOEFL Sections on Listening Comprehension and on Structure and Written Expression remain fairly stable from the first test administration to the second test administration, there is a sizeable change in the average

correlation for the TOEFL Section on Reading Comprehension and Vocabulary (i.e., $\bar{r}$ = .68 for Administration I and $\bar{r}$ = .83 for Administration II). Almost every entry in the correlation matrix for Section III in Administration II is higher than the corresponding entry in Administration I (see Tables 2 and 3). The likelihood that this merely reflects changes in the composition of the samples drawn for the respective language groups from one test administration to another seems low given the greater stability of the average correlations for the other two TOEFL Sections (i.e., $\bar{r}$ = .83 and .78 on Section I, and $\bar{r}$ = .78 and .81 on Section II). It seems more probable that the change in the average correlation for the TOEFL Section on Reading Comprehension and Vocabulary is attributable to variation in test forms. This finding is especially important because Swinton and Powers (1980) based their exploratory factor analysis of the TOEFL on the same data set from Administration I, the form with lower entries in its correlation matrix of item difficulty values across language groups for the TOEFL Section on Reading Comprehension and Vocabulary, and found evidence suggesting a separate factor for vocabulary. The variation in test forms on the pertinent TOEFL Section observed here suggests that further research on the test's factor structure should be undertaken in order to verify the existence of a separate vocabulary dimension of English proficiency and that particular attention should be given to the test specifications for this section in developing equivalent forms.

Differential Item Performance Across Language Groups

Detecting instances in which the item performance of examinees of comparable ability differed by language groups entailed contrasting the item performance of several language groups at the same score level.

Accordingly, the score distribution of the complete sample of language groups was divided into ten score intervals for each section of the test, with approximately equal numbers of examinees at each score level. It was also required that each language group have at least ten examinees at any particular score stratum as a check for an arbitrary minimum frequency in the marginal distribution of language groups by score strata, $n_{i+j} \geq 10$. These score distributions by language groups appear in Appendix C for each test section and both test administrations.

The Germanic language group so far exceeded the other language groups in their test performance on each TOEFL Section (see Table 1 and Appendix C) that it was not possible to include seven language groups and ten score strata in calculating chi-square values and still maintain minimum frequencies in the marginal distributions of language groups by score strata. Fewer than one-tenth of the Germanic examinees were in the lower half of the score distribution for the TOEFL Section on Listening Comprehension and fewer than ten Germanic examinees were always in the lowest tenth of the score distributions for the other two TOEFL Sections. Consequently, the Germanic language group was omitted from the chi-square analysis for the TOEFL Section on Listening Comprehension and the ninth and tenth score strata were collapsed together in the chi-square analyses for the TOEFL Sections on Structure and Written Expression and on Reading Comprehension and Vocabulary.

The number of items on which the chi-square analysis resulted in a statistically significant value is given in Table 4. Clearly the proportion of items shown to elicit differential performance across various language groups, nearly seven-eighths of the total number of items, is greater than the proportions identified through statistical indices in other

Table 4

Identification of Discrepant Item Performance by Chi-Square Index

Test Section	Administration I		Administration II	
	Significant	Not Significant	Significant	Not Significant
I. Listening Comprehension	38[a]	12	40[a]	10
II. Structure and Written Expression	37[b]	2	36[b]	4
III. Reading Comprehension and Vocabulary	57[b]	3	53[b]	7

a χ^2 (50) > 86.66, $p < .001$

b χ^2 (54) > 91.84, $p < .001$

studies. This is probably attributable to the natural differences among the language groups in the extent of the similarities and dissimilarities of their languages with the English language. That such contrasts played a role in the results was particularly evident with respect to vocabulary: the highest chi-square values were obtained for items from the TOEFL Section on Reading Comprehension and Vocabulary, and these were items requiring examinees to choose a synonym for a given word in a sentence. It is perhaps also attributable, in part, to the number of groups, number of score strata, and size of samples employed in the chi-square analysis for this study.

Percentage deviations from expected frequencies of correct item responses begin to explain statistical significance in this context. Table 5 presents chi-square values for the items from the TOEFL Section on Reading Comprehension and Vocabulary in Administration I, the section which had the lowest intercorrelations of item difficulty values between language groups, along with each group's relative advantage or disadvantage on these items expressed as a percentage deviation from the expected frequency of correct item responses for that particular group and item. (Appendix D gives the total chi-square values, percentage deviations, and item difficulty values for all items.) The extent of these deviations from expected performance tends to follow the magnitude of the chi-square values for the items. A significant chi-square value, however, may result from very modest variations in item performance across groups. Item 45 had a significant chi-square value, $\chi^2_{45}(54) = 92.39$ and $p < .001$, but showed just a seven percent increase in correct responses among Chinese examinees and a six percent decrease among Spanish examinees. Item 14, on the other hand, had a much higher chi-square value, $\chi^2_{14} = 1408.41$ and $p < .001$, and a

Table 5

Percentage Deviations from Expected Frequencies of Correct Item Responses

Section III: Reading Comprehension and Vocabulary

Administration I

Item Number	Total Chi-square	Interaction Chi square	Percentage Deviations from Expected Frequencies of Correct Item Responses						
			Germanic	Spanish	African	Chinese 1	Chinese 2	Japanese	Arabic
14	1408	94	12	36	40	-33	-32	-26	-2
22	1379	101	-7	27	18	-38	-35	48	-27
23	755	106	-26	27	67	-5	-14	-39	-18
19	739	97	1	22	32	-32	-35	17	-8
25	653	182	-7	41	-5	-14	-19	-5	-12
21	648	56	16	-43	26	25	23	-12	-24
26	627	119	-9	-7	35	22	25	-42	-32
24	569	83	-40	42	2	-5	-7	-2	13
27	556	91	-24	-35	31	4	8	41	-29
18	542	98	-6	-18	25	2	-3	-23	48
12	540	104	10	-24	11	17	20	-4	7
59	516	97	7	5	-38	5	6	8	0
6	512	60	7	7	-20	-9	-11	10	19
1	499	57	7	11	-13	-13	-15	15	12
10	481	79	6	-1	34	-21	-21	0	16
15	441	154	-16	11	60	-36	-29	6	16
16	433	46	-16	-1	46	16	3	-22	-34
33	427	115	19	-12	-22	19	26	-19	-14
58	412	64	10	-27	0	14	21	-2	-7
28	375	78	16	27	22	-29	-31	-5	-22
13	375	45	-8	16	17	-3	1	-15	-14
2	362	58	3	-12	7	3	3	-1	0
51	357	70	2	-5	-27	20	20	-13	1
50	344	105	7	-2	-25	12	11	-6	-1
7	338	116	-1	-4	-8	-9	-11	21	13
8	314	81	-7	17	1	-7	-9	5	-6
53	309	74	5	-1	-21	15	13	-18	6
60	303	96	3	-11	-14	9	11	7	-4
36	273	68	9	-9	-17	11	10	0	-4
49	271	95	1	-7	-2	11	9	-13	2
30	265	171	-44	10	55	-5	-5	-13	3
46	261	78	1	-9	2	6	6	-2	-3

Item Number	Total Chi-square	Interaction Chi-square	Percentage Deviations from Expected Frequencies of Correct Item Responses						
			Germanic	Spanish	African	Chinese 1	Chinese 2	Japanese	Arabic
55	258	81	0	-4	-12	9	9	-4	-1
39	258	42	4	-3	-16	-3	0	16	0
4	253	95	9	6	-11	-9	-9	10	5
54	250	89	0	-3	-11	11	8	-1	-7
3	220	94	-3	-12	5	8	8	-6	4
52	214	102	-3	-12	-5	17	14	-13	11
17	212	100	-3	17	-21	-4	-6	14	-8
35	195	36	1	-1	-9	2	2	6	-4
57	184	55	14	4	-24	-4	-9	16	-2
9	183	45	8	-12	3	4	3	-8	10
48	180	63	-5	10	-19	-5	1	19	-13
5	174	42	6	12	-1	-10	-7	4	-9
42	172	56	5	7	-8	-7	-5	2	7
37	152	55	3	-2	-3	4	6	-11	4
56	141	59	4	-3	-13	3	6	3	-2
34	137	72	3	-1	-7	1	0	4	-1
20	137	63	-12	-12	0	9	13	-3	8
11	135	69	-5	-8	8	-2	-4	9	6
29	130	75	6	9	19	-8	-12	-10	-10
38	122	57	-8	-6	3	5	4	-1	5
41	117	64	4	-4	-6	-2	-2	5	9
31	106	61	4	4	-1	0	-3	-2	-3
43	102	40	-4	-5	2	3	5	2	-6
40	93	64	5	3	-2	-1	2	-4	-3
45	92	69	-2	-6	0	7	5	-3	-1
32	88	50	1	1	2	0	-1	-5	2
44	81	46	3	-5	-2	5	3	-3	-1
47	80	44	-1	-2	7	2	-1	-8	7

much greater range in percentage deviations among groups. Statistical significance by itself can vary widely in its implications for the relative advantages and disadvantages held by different groups on a test item.

Both the chi-square procedure and the calculation of percentage deviations assume that differential item performance by groups holds across ability levels. Otherwise the aggregation across score strata might obscure differences from level to level in item discrimination and distort relative advantages and disadvantages among groups. The second column of chi-square values given in Table 5 reveal the significance of the three-way interaction among item responses, language groups, and score strata for each item. For half of the items this interaction was not statistically significant, $X_2(48) < 73.67$ and $p > .01$, and the likelihood-ratio chi-square values for the interaction even when statistically significant were much smaller than the overall chi-square values. The relative advantages and disadvantages for different language groups reported as percentage deviations from expected frequencies of correct item responses in Table 5 seem to be fairly stable across strata in the score distribution.

The distributions of percentage deviations by language groups appear in Table 6 for the TOEFL Sections in Administration I and in Table 7 for the TOEFL Sections in Administration II. These distributions illustrate the consistency of item performance by language groups, given the item performance of the total sample as a standard for comparison, and the frequency of relative advantages and disadvantages by language groups. Each distribution necessarily centers near zero because scores on the test sections served as an internal criterion for expected frequencies of correct item responses and, in effect, precluded skewed distributions of percentage

Table 6
Distribution of Percentage Deviations from Expected Frequencies of Correct Item Responses by Language Groups: Administration I

Test Section	Percentage Deviation	Language Groups						
		Germanic	Spanish	African	Chinese 1	Chinese 2	Japanese	Arabic
Section I: Listening Comprehension	> +50							
	+40 to +50			1				
	+30 to +40			3				
	+20 to +30			2			4	
	+10 to +20		3	6	4	3	3	8
	0 to +10		21	14	24	27	16	15
	-10 to 0		22	13	16	15	18	18
	-20 to -10		4	6	5	3	5	9
	-30 to -20			3	1	2	3	
	-40 to -30			1				
	-50 to -40			1			1	
Section II: Structure and Written Expression	> +50					1		
	+40 to +50		1	1	1			
	+30 to +40		1		1	1		
	+20 to +30	1	2	5	1		4	
	+10 to +20	4	4	5	5	4	5	4
	0 to +10	21	13	9	12	13	12	16
	-10 to 0	7	8	11	12	13	10	12
	-20 to -10	1	7	2	4	2	4	4
	-30 to -20	2	1	6	3	5	1	2
	-40 to -30	3	1				1	1
	-50 to -40		1					
	< -50		1				2	
Section III: Reading Comprehension and Vocabulary	> +50			3				
	+40 to +50		2	2			2	1
	+30 to +40		1	4				
	+20 to +30		4	3	3	6	1	
	+10 to +20	6	8	3	9	6	8	8
	0 to +10	28	9	12	21	20	12	17
	-10 to 0	19	24	16	18	17	22	22
	-20 to -10	3	8	9	3	5	10	6
	-30 to -20	2	2	7	2	2	3	2
	-40 to -30		1	1	4	4	1	
	-50 to -40	2	1				1	

Table 7
Distribution of Percentage Deviations from Expected Frequencies of Correct Item Responses by Language Groups: Administration II

Test Section	Percentage Deviation	Germanic	Spanish	African	Chinese 1	Chinese 2	Japanese	Arabic
Section I: Listening Comprehension	> +50			1				
	+40 to +50			1			1	
	+30 to +40			4				
	+20 to +30			4			1	3
	+10 to +20		6	4	4	4	9	7
	0 to +10		22	11	21	20	12	15
	-10 to 0		16	11	20	20	18	15
	-20 to -10		6	5	4	3	5	7
	-30 to -20			7	1	3	4	3
	-40 to -30			1				
	-50 to -40			1				
Section II: Structure and Written Expression	> +50			1				
	+40 to +50							
	+30 to +40		2	1				
	+20 to +30		2	1	3	2		
	+10 to +20	3	4	5	2	4	6	5
	0 to +10	21	7	16	16	14	14	16
	-10 to 0	14	18	6	14	17	13	13
	-20 to -10	1	6	7	2		5	4
	-30 to -20	1		3	1	1	1	2
	-40 to -30		1				1	
	-50 to -40				2	2		
Section III: Reading Comprehension and Vocabulary	> +50		1	1				
	+40 to +50			2				
	+30 to +40		3	4	1			
	+20 to +30	1	2	6		1	4	2
	+10 to +20	10	6	5	6	6	6	7
	0 to +10	27	21	9	25	27	22	19
	-10 to 0	12	16	19	20	16	14	18
	-20 to -10	4	5	5	5	8	7	9
	-30 to -20	3	3	4	3	2	5	5
	-40 to -30	2	1	1			2	
	-50 to -40	1	2	1				
	< -50			3				

deviations indicating an overall advantage or disadvantage on the test for a particular group. The modal deviation for each distribution is within ten percent of the expected frequency of correct item responses, and the majority of deviations occur within this same range except for the African language group on TOEFL Sections II and III. Large deviations from expected item performance (i.e., over thirty percent greater or fewer correct item responses than expected) are infrequent but occur more often in the TOEFL Section on Reading Comprehension and Vocabulary, perhaps suggesting that knowledge of specific words in the context of single sentences and reading passages is more susceptible to linguistic contrasts than either aural skills or syntactical rules. The extent and spread of percentage deviations within the African group further suggests the heterogeneous nature of this group, which included four different native languages (i.e., Ibo, Yoruba, Efik, and Fanti) more on the basis of their geographical proximity than their linguistic comparability.

Linguistic Explanations for Discrepant Item Performance

By expressing each group's relative advantage or disadvantage on an item, percentage deviations can facilitate examining the possible reasons for discrepant item performance among groups. Item 14 from the TOEFL Section on Vocabulary and Reading Comprehension in Administration I provides a convenient example since it had the highest chi-square value of any item. That item required examinees to choose the response which best kept the meaning of a word in a sentence:

14. The card had perforations in the upper left corner.

(A) holes (B) numbers (C) symbols (D) lines

Just the percentages of correct responses on this item revealed substantial variation across language groups: 37% and 38% correct for the two Chinese samples; 40% correct for the Japanese sample; 50% correct for the Arabic sample; 90% correct for both the African and Germanic samples; and 99% correct for the Spanish sample. The percentage deviations from expected frequencies of correct responses for the item (see Table 5) showed a somewhat different pattern of relative advantage and disadvantage among the language groups. The Chinese and Japanese examinees indeed had fewer correct responses than expected, -33% and -26%, primarily because these language groups tended to choose the distractor "symbols" which perhaps seemed to fit the content of the item stem. Despite attaining the same proportion of correct responses, African examinees found this item easier than other items in the section, a +40% deviation from their expected item performance, while Germanic examinees found the item only somewhat easier, a +12% deviation. The performance and advantage of Spanish examinees on this item undoubtedly stems from cognates for both "holes" and "perforations" in their native language, "hoyo" and "perferacion." Similarly, Germanic examinees probably benefited from the noun "höhle" and the verb "perforieren" in their own language. The performance of Arabic examinees on this item was consistent with their performance on other items in the same test section as reflected by the slight percentage deviation from the expected frequency of correct responses for this language group. Percentage deviations from expected frequencies of correct item responses, unlike simple proportions or percentages of correct item responses, take score into account and therefore seem to yield richer information for the interpretation of differential item performance across groups.

Specialists in English as a foreign language reviewed the results of the analyses for Administration I and attempted to identify probable instances of discrepant item performance given only the test form and answer key for Administration II. The complexity of the reviewers' task became readily apparent. Seven reviewers examined the most extreme cases of differential item performance found in the first test administration and offered possible explanations for the observed differences among language groups. But these explanations necessarily depended on each reviewer's familiarity with the several native languages and the linguistic similarities and dissimilarities with the sound, syntax, semantics, and vocabulary of the English language. Usually a reviewer was familiar with two or three languages and offered comments about an item based only on the performance of those particular language groups.

The success of the reviewers in identifying instances of discrepant item performance on a second test form demonstrate the difficulty of judgmental approaches to this problem. Six reviewers made informed guesses, or "shots in the dark" as one specialist described them, about the ten items from each test section which might exhibit the greatest differential performance across language groups and thus result in the highest chi-square values. Asked to identify the ten most discrepant items of the fifty items in the TOEFL Section on Listening Comprehension, the reviewers were correct 12 times in 64 tries. Asked to identify the ten most discrepant items of the forty items in the TOEFL Section on Structure and Written Expression, the reviewers were correct 13 times in 53 tries. Both of these success ratios are simply at the chance level. Asked to identify the ten most discrepant items of the sixty items in the TOEFL

Section on Reading Comprehension and Vocabulary, the reviewers were correct 19 times in 60 tries--a success ratio close to double the chance level but still too unreliable for practical application. One vocabulary item, which ranked third in the order of chi-square values for items within this section, was chosen by all six reviewers: the stem and the correct response on this item were both Spanish cognates and the Spanish language group did show a sizeable relative advantage on the item. Five reviewers also chose another specific vocabulary item; the observed frequencies of correct responses on that item were within ten percent of the expected frequencies for each language group.

A seventh specialist had declined to participate in the second phase of the reviewers' task, noting that a priori contrastive analysis had not proven to be effective in its application, at least to nonphonological areas, and suggesting several references. Contrastive analysis calls for systematic comparisons of the native language and the second language with prescriptions for teaching based on these comparisons. Wardhaugh (1970) noted that such contrastive analysis "makes demands of linguistic theory, and, therefore, of linguists, that they are in no position to meet" (p. 125). An a posteriori analysis of observed difficulties in second language acquisition may become helpful in teaching and testing, but a priori contrastive analysis reliant on linguistic similarities and differences alone may be just speculation (Schachter & Celce-Murcia, 1977; Wardhaugh, 1970).

Judgmental approaches to identifying items which might exhibit differential performance among language groups depend too heavily on simultaneous a priori contrastive analyses of several different languages with the English language. Detecting instances of differential item

performance by language groups would seem to require the selection and application of both an appropriate statistical procedure and a meaningful measure of each group's relative advantage or disadvantage on an item.

SUMMARY AND CONCLUSIONS

A high proportion of TOEFL items, nearly seven-eighths of those involved in this study, were found to be sensitive to examinees' native languages. Examinees from several language groups, despite comparable test scores, differed in their performance on specific items according to their native language. Specialists in English as a second language attributed such differential item performance to linguistic similarities and dissimilarities with the English language. Yet these same reviewers were unable to predict which items would exhibit marked discrepancies in performance across language groups from a test form and answer key alone.

It seems obvious that a native language influences acquisition and performance in a second language. Less clear and perhaps unknown is the degree to which a test of proficiency in a second language accurately reflects linguistic affinity with an examinee's native language. Vocabulary test items, for example, may happen to include cognates of a particular language as both stem and key and thus offer examinees from that language group a distinct advantage. Statistical procedures now available can isolate instances in which performance on a given item is inconsistent across groups after overall ability has been taken into consideration.

But a significant result from a statistical procedure for detecting discrepant item performance seems to constitute insufficient grounds for rejecting an item in this context. There were cases here in which deviations of less than ten percent from the expected frequencies of correct

item responses across language groups led to a statistically significant index of discrepant performance on an item. In other cases the percentage deviations from expected frequencies of correct item responses seemed to reflect contrasts with the English language. Still other items, however, may lead to exaggerated or unexplained deviations from the expected item performance across language groups, performance inconsistent with examinees' ability and unaccountable by linguistic similarities and dissimilarities. These latter items could distort examinees' scores across language groups as well as across test forms.

The statistical procedure for detecting discrepant item performance followed in this study was a chi-square analysis. Unlike other procedures, it lends itself readily to the examination of multiple groups and multiple dimensions. Here the concern was comparisons of item performance across several language groups; other studies might focus on individual items and patterns of response choices, including distractors, across groups or expand the number of dimensions so as to investigate differences in item performance by sex or level of education. The chi-square procedure can also provide measures of item discrepancy for each group, such as percentage deviations from expected frequencies of correct item responses, and lead naturally to estimates of other effects or interactions through log-linear models. Moreover, this statistical procedure is practical for routine use in developing tests, in that its application is straightforward and efficient. But like other statistical procedures reliant on an internal ability criterion, chi-square analysis is effective in detecting instances of discrepant item performance only insofar as the test score itself represents a fair measure of the construct of interest.

REFERENCES

Angoff, W. H. The use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore, Md.: Johns Hopkins University Press, in press.

Angoff, W. H., & Ford, S. F. Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 1973, 10, 95-106.

Angoff, W. H., & Sharon, A. T. The evaluation of differences in test performance of two or more groups. Educational and Psychological Measurement, 1974, 34, 807-816. [See also Angoff, W. H., & Sharon, A. T. Patterns of test and item difficulty for six foreign language groups on the Test of English as a Foreign Language (ETS RB-72-2). Princeton, N.J.: Educational Testing Service, 1972.]

Baker, F. B. A criticism of Scheuneman's item bias technique. Journal of Educational Measurement, 1981, 18, 59-62.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. Discrete multivariate analysis: Theory and practice. Cambridge, Mass.: MIT Press, 1975.

Cleary, T. A., & Hilton, T. L. An investigation of item bias. Educational and Psychological Measurement, 1968, 28, 61-75.

Ironson, G. H., & Subkoviak, M. J. A comparison of several methods of assessing item bias. Journal of Educational Measurement, 1979, 16, 209-225.

Lord, F. M. A study of item bias, using item characteristic curve theory. In Y. H. Poortinga (Ed.), Basic problems in cross-cultural research. Amsterdam: Swets and Zeitlinger, 1977, pp. 19-29.

Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, N.J.: Erlbaum, 1980.

Rudner, L. M., Getson, P. R., & Knight, D. L. A Monte Carlo comparison of seven biased item detection techniques. Journal of Educational Measurement, 1980, 17, 1-10. (a)

Rudner, L. M., Getson, P. R., & Knight, D. L. Biased item detection techniques. Journal of Educational Statistics, 1980, 5, 213-233. (b)

Schachter, J., & Celce-Murcia, M. Some reservations concerning error analysis. TESOL Quarterly, 1977, 11, 441-451.

Scheuneman, J. A method of assessing bias in test items. Journal of Educational Measurement, 1979, 16, 143-152.

Scheuneman, J. D. A response to Baker's criticism. Journal of Educational Measurement, 1981, 18, 63-66.

Scheuneman, J. D. A posteriori analyses of biased items. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore, Md.: Johns Hopkins University Press, in press.

Shepard, L., Camilli, G., & Averill, M. Comparison of six procedures for detecting test item bias using both internal and external ability criteria. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, 1980.

Sinnott, L. T. Differences in item performance across groups (ETS RR-80-19). Princeton, N.J.: Educational Testing Service, 1980.

Stricker, L. J. A new index of differential subgroup performance: Application to the GRE Aptitude Test (ETS Research Report, RR-81-13). Princeton, N.J.: Educational Testing Service, 1981.

Swinton, S. S., & Powers, D. E. Factor analysis of the Test of English as a Foreign Language for several language groups (TOEFL Research Report 6). Princeton, N.J.: Educational Testing Service, 1980.

Wardhaugh, R. The contrastive-analysis hypothesis. TESOL Quarterly, 1970, 4, 123-130.

Appendix A

Descriptive Profiles of Language Groups

Descriptive Profile of Language Groups:
Administration I

	Language Groups													
	Germanic		Spanish		African		Chinese 1		Chinese 2		Japanese		Arabic	
	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
Sex														
Male	371	62	733	74	641	87	624	63	643	64	705	71	613	89
Female	225	38	258	26	99	13	374	37	355	36	292	29	73	11
Reason for TOEFL														
Undergraduate application	228	38	317	32	620	84	617	62	628	63	321	32	279	41
Graduate application	297	50	564	57	98	13	211	21	218	22	512	51	262	38
Other school application	11	2	12	1	8	1	112	11	88	9	17	2	5	1
Professional licensure	21	3	40	4	1	0	11	1	15	2	34	3	53	8
Business requirement	15	2	25	3	1	0	20	2	24	2	66	7	74	11
Other	22	4	30	3	1	0	21	2	20	2	37	4	9	1
Previous TOEFL administrations														
0	582	98	852	86	696	94	731	73	729	73	577	58	534	78
1	14	2	117	12	41	6	205	21	208	21	278	28	87	13
2			16	2	3	0	43	4	43	4	85	9	36	5
3			5	0			14	1	14	1	32	3	28	4
4 or more			1	0			5	1	4	0	25	2	1	0
Native Country														
Germany	246	41												
Netherlands	112	20												
Sweden	74	12												
Switzerland	30	5												
Mexico			214	22										
Columbia			113	11										
Venezuala			82	8										
Peru			71	7										
Argentina			69	7										
Chile			59	6										
Spain			49	5										
Nigeria					688	93								
Ghana					47	6								
Hong Kong							846	85	850	85				
China							52	5	57	6				
Malaysia							50	5	41	4				
Japan											976	98		
Jordan													213	31
Egypt													126	18
Saudi Arabia													116	17
Lebanon													81	12
Iraq													35	5

Descriptive Profile of Language Groups:

Administration II

	Language Groups													
	Germanic		Spanish		African		Chinese 1		Chinese 2		Japanese		Arabic	
	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
Sex														
Male	501	52	667	67	397	88	597	60	601	60	621	62	867	88
Female	459	48	325	33	52	12	396	40	395	40	375	38	120	12
Reason for TOEFL														
Undergraduate application	543	57	426	43	383	85	282	28	282	28	402	40	507	51
Graduate application	323	34	480	48	41	9	633	64	650	65	414	42	390	39
Other school application	15	2	22	2	6	1	43	4	31	3	31	3	10	1
Professional licensure	5	1	17	2	6	1	8	1	5	1	18	2	34	3
Business requirement	24	2	26	3	1	0	6	1	9	1	74	7	13	1
Other	44	5	14	1	1	0	6	1	7	1	45	4	8	1
Previous TOEFL administrations														
0	912	95	764	77	401	89	601	60	570	57	420	42	728	74
1	44	5	162	16	41	9	244	25	276	28	291	29	165	17
2	3	0	44	4	6	1	82	8	95	10	136	14	63	6
3	1	0	11	1	1	0	35	4	32	3	67	7	18	2
4 or more	0	0	11	1	0	0	31	3	23	2	82	8	13	1
Native Country														
Germany	516	54												
Netherlands	150	16												
Sweden	100	10												
Denmark	66	7												
Mexico			164	16										
Venezuala			148	15										
Columbia			144	14										
Argentina			82	8										
Peru			74	7										
Panama			59	6										
Spain			50	5										
Nigeria					397	89								
Ghana					42	9								
China							677	68	689	69				
Hong Kong							283	29	276	28				
Japan											978	98		
Lebanon													292	30
Jordan													206	21
Egypt													125	13
Saudi Arabia													82	8

Descriptive Profile of Language Groups:

Years of Age

	Administration I		Administration II	
	Mean	sd	Mean	sd
Germanic	23.14	4.3	22.09	3.7
Spanish	23.89	5.4	23.01	5.9
African	23.65	5.2	23.40	4.7
Chinese 1	20.22	3.1	24.13	4.5
Chinese 2	20.39	3.3	24.20	4.7
Japanese	25.46	5.1	24.27	4.9
Arabic	24.02	5.8	22.96	5.6

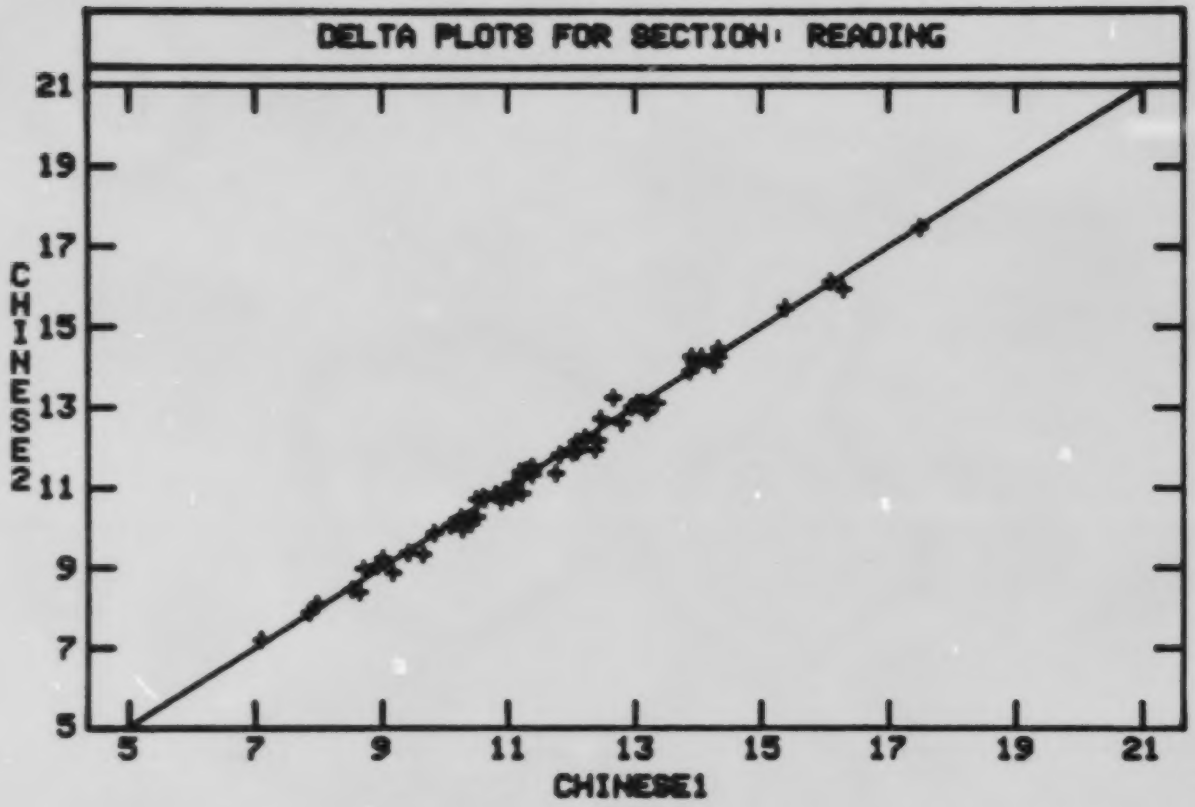
Appendix B

Illustrative Delta Plots of Item Difficulty Values

TOEFL Section III: Reading Comprehension and Vocabulary

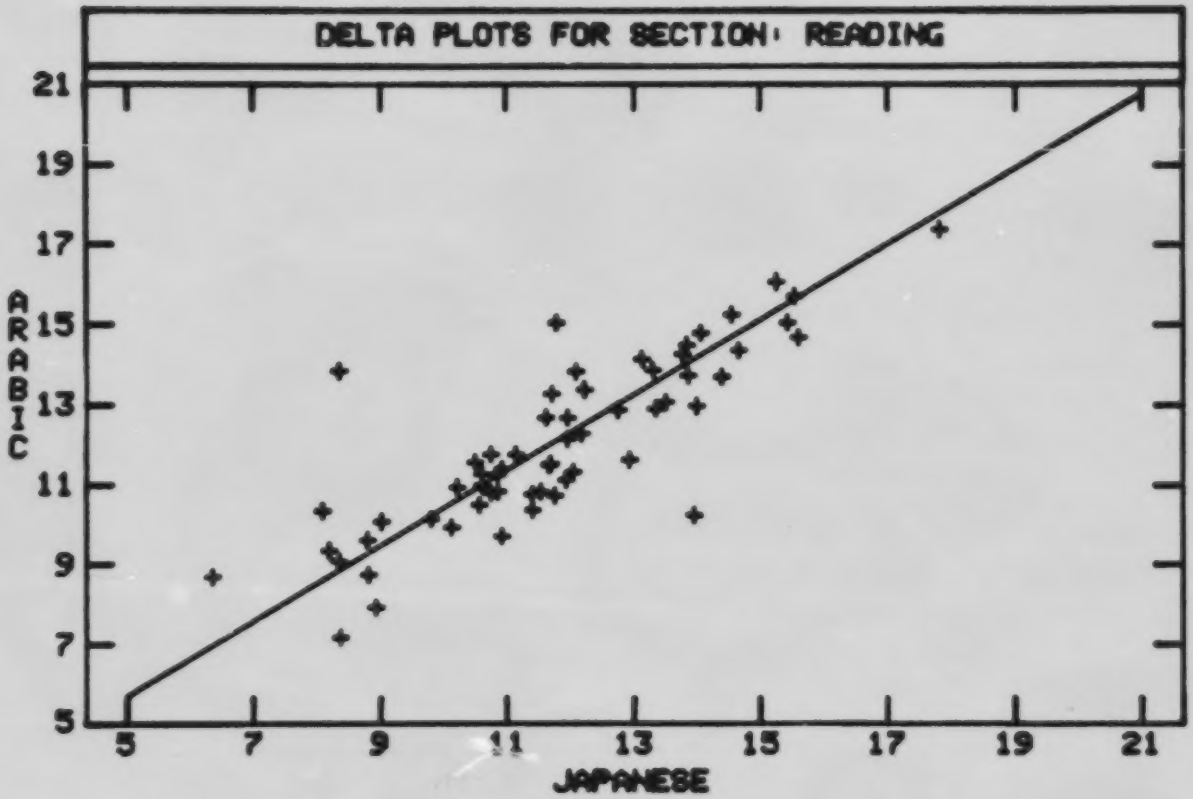
Administration I

CHINESE 1 & CHINESE 2 ADMIN 1



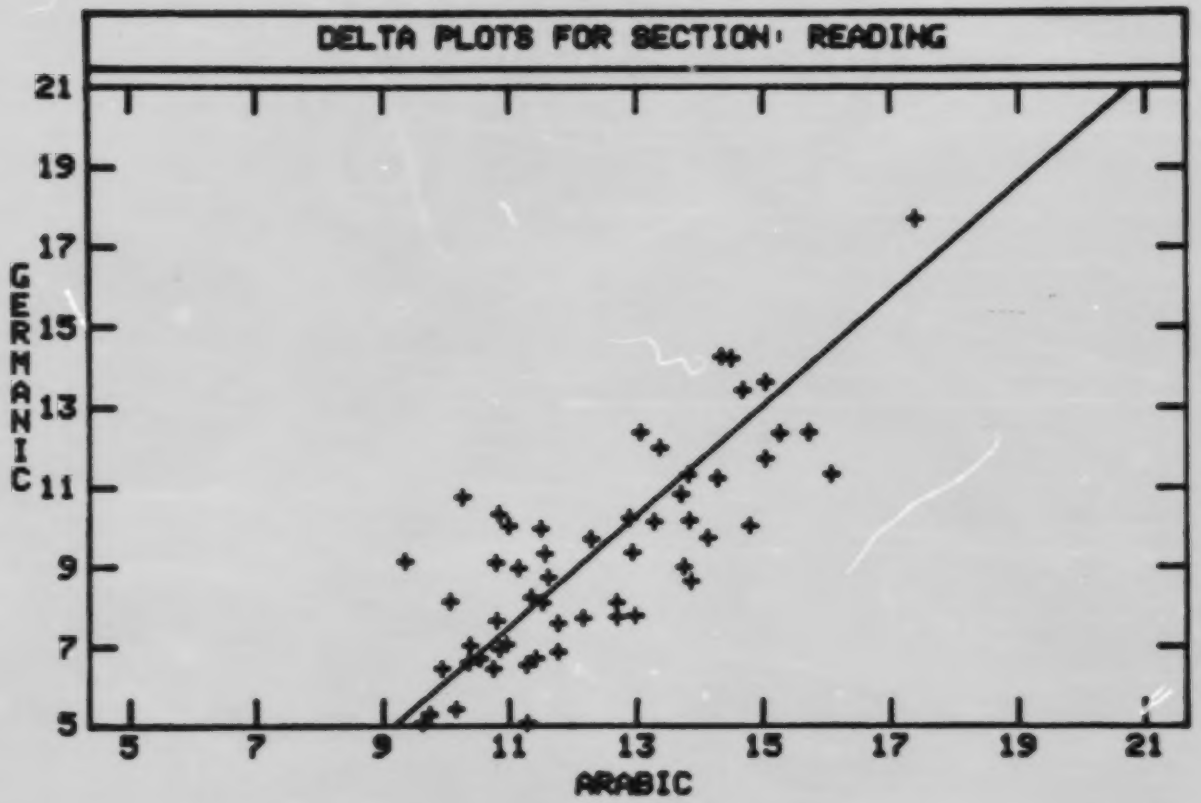
r = .99+

JAPANESE & ARABIC ADMIN 1



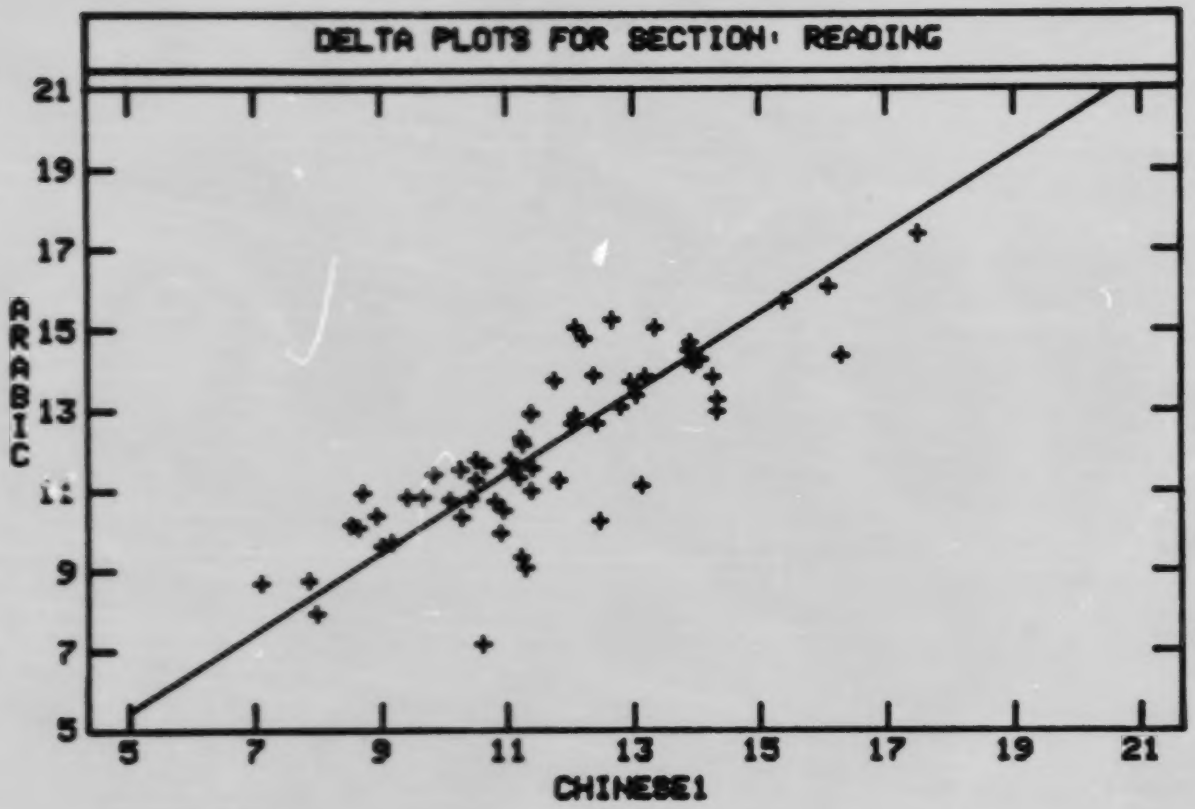
r = .84

ARABIC & GERMANIC ADMIN 1



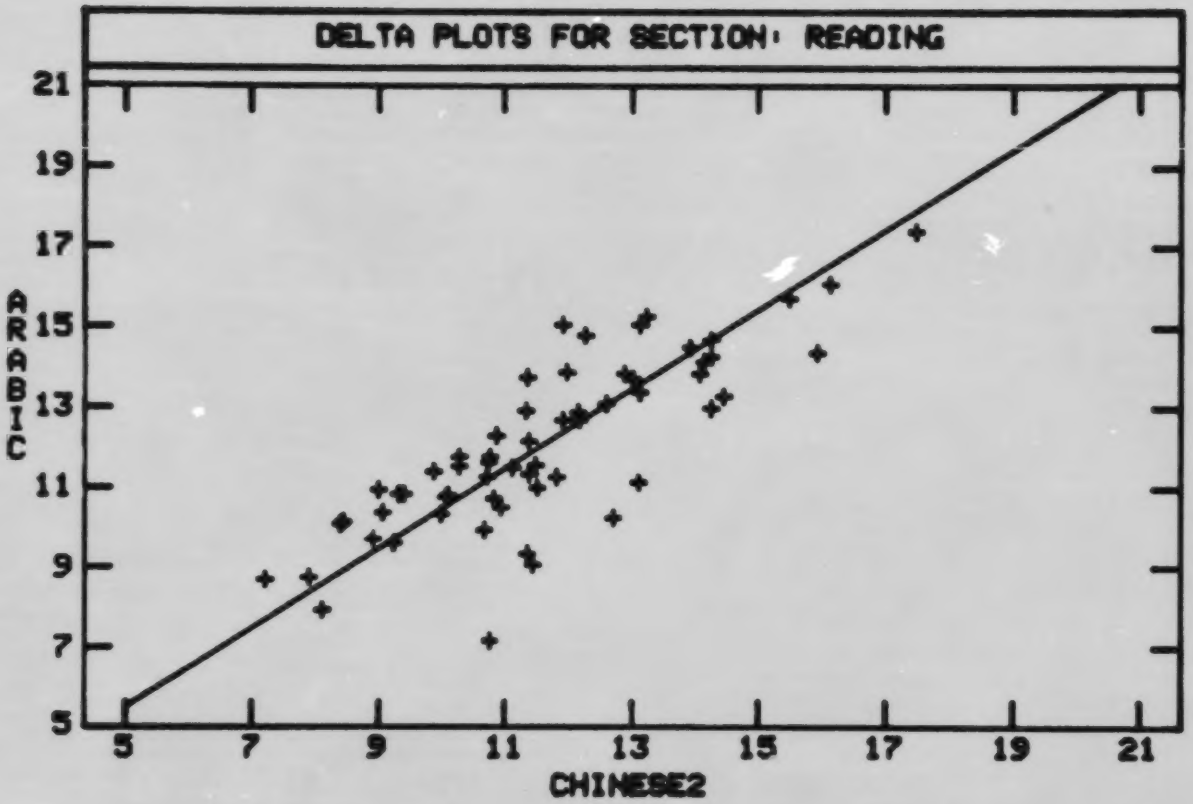
r = .83

CHINESE 1 & ARABIC ADMIN I



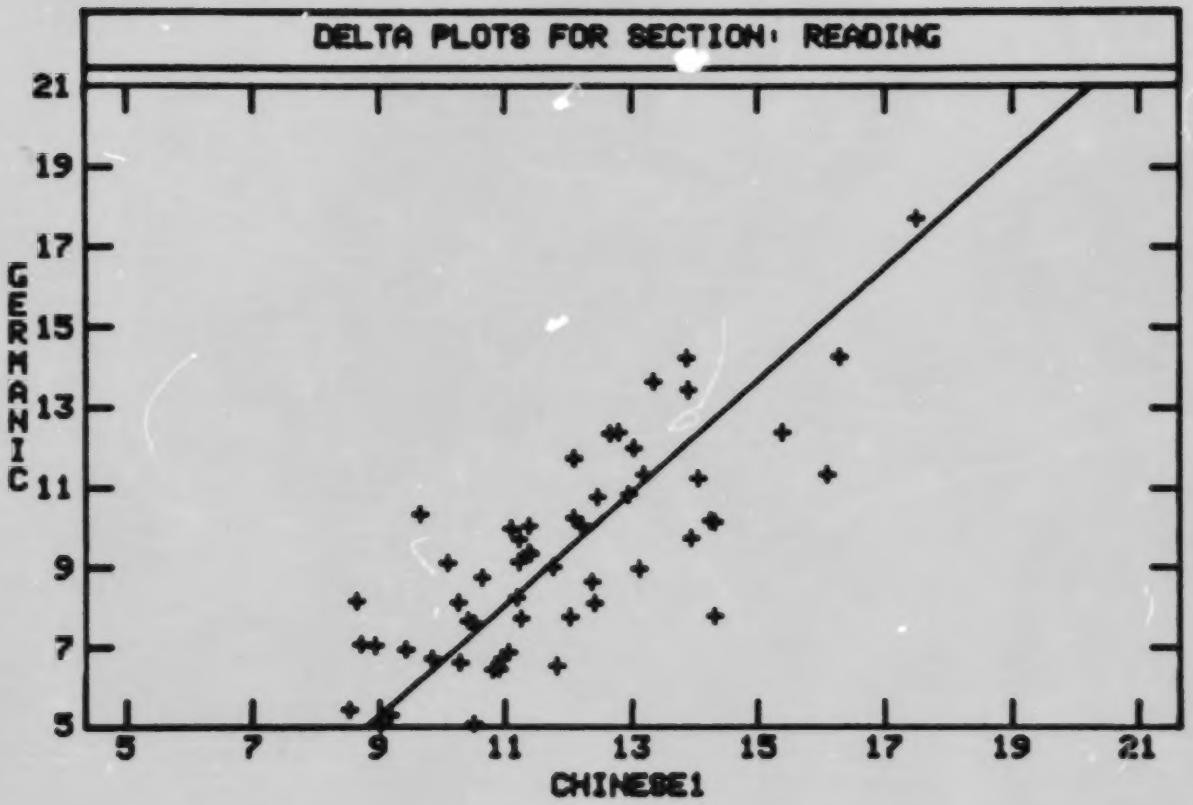
r = .83

CHIESE 2 & ARABIC ADMIN 1



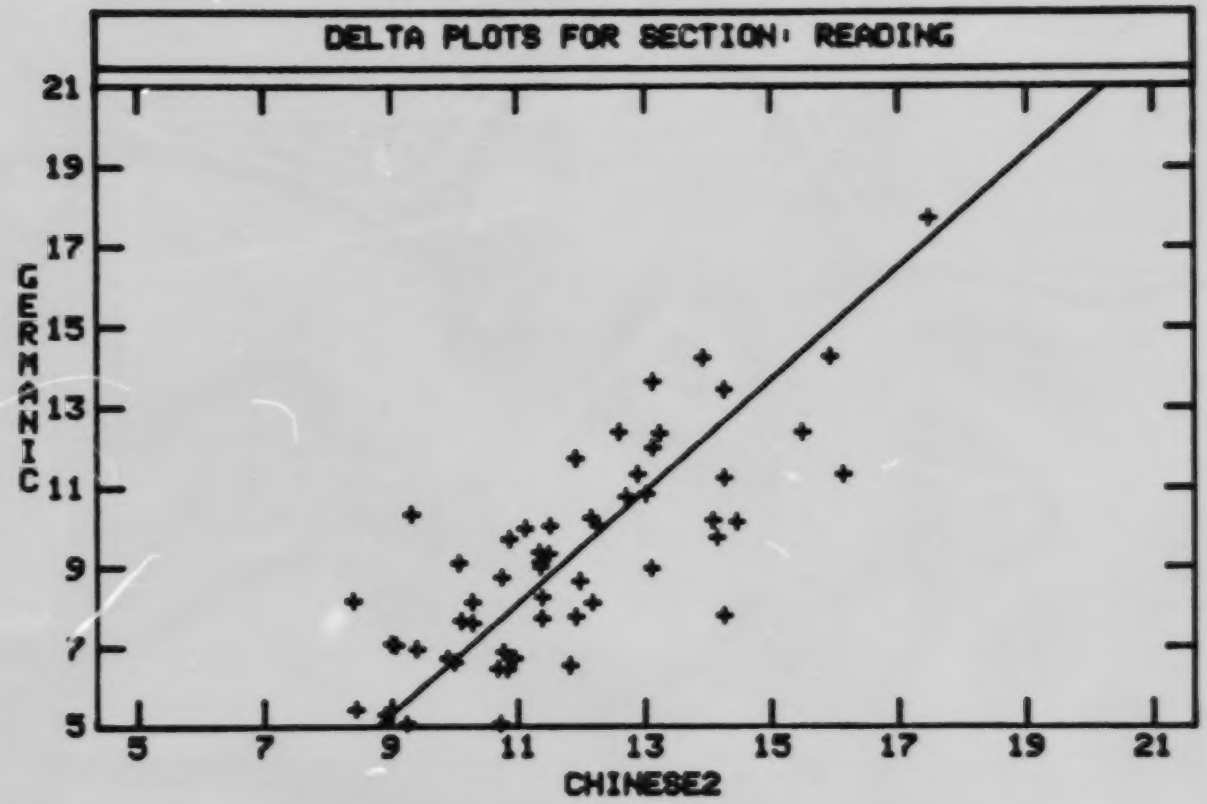
r = .82

CHINESE 1 & GERMANIC ADMIN 1

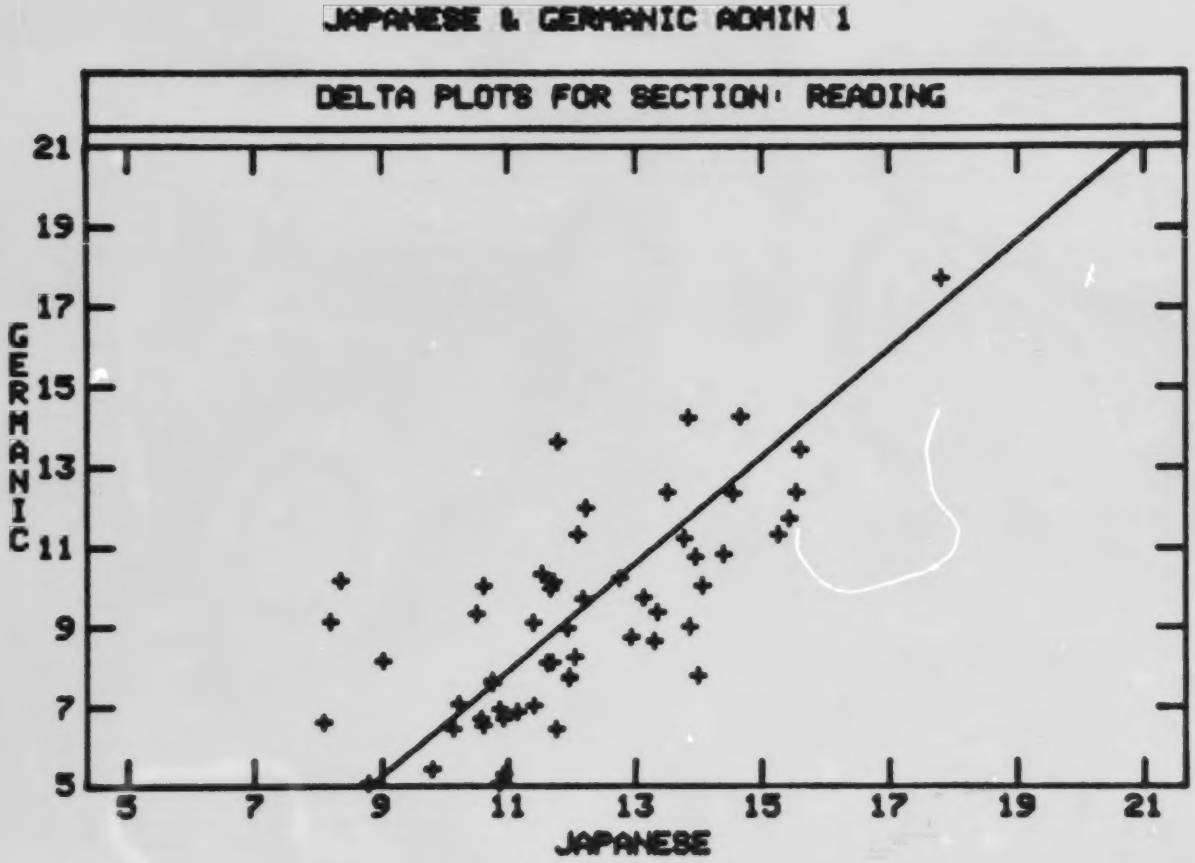


r = .80

CHINESE 2 & GERMANIC ADMIN 1

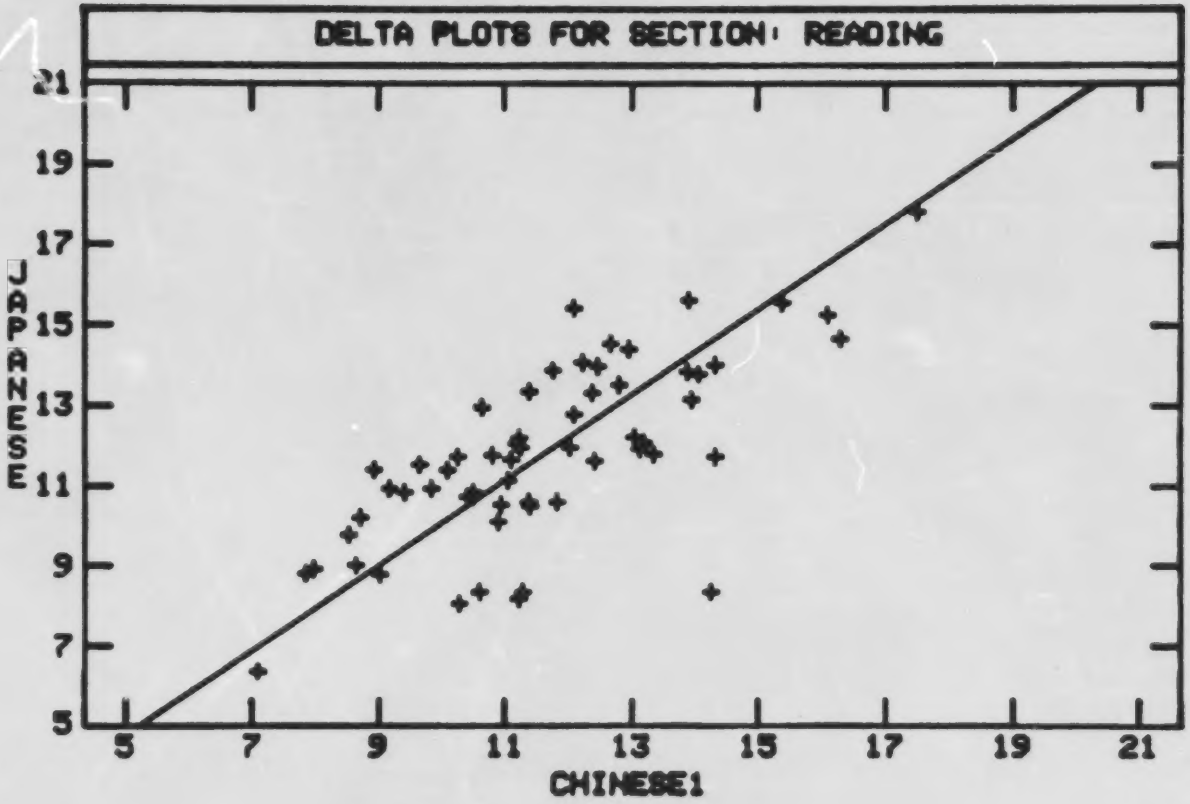


r = .79


JAPANESE & GERMANIC ADMIN 1
DELTA PLOTS FOR SECTION: READING
GERMANIC
JAPANESE
21
19
17
15
13
11
9
7
5
5
7
9
11
13
15
17
19
21

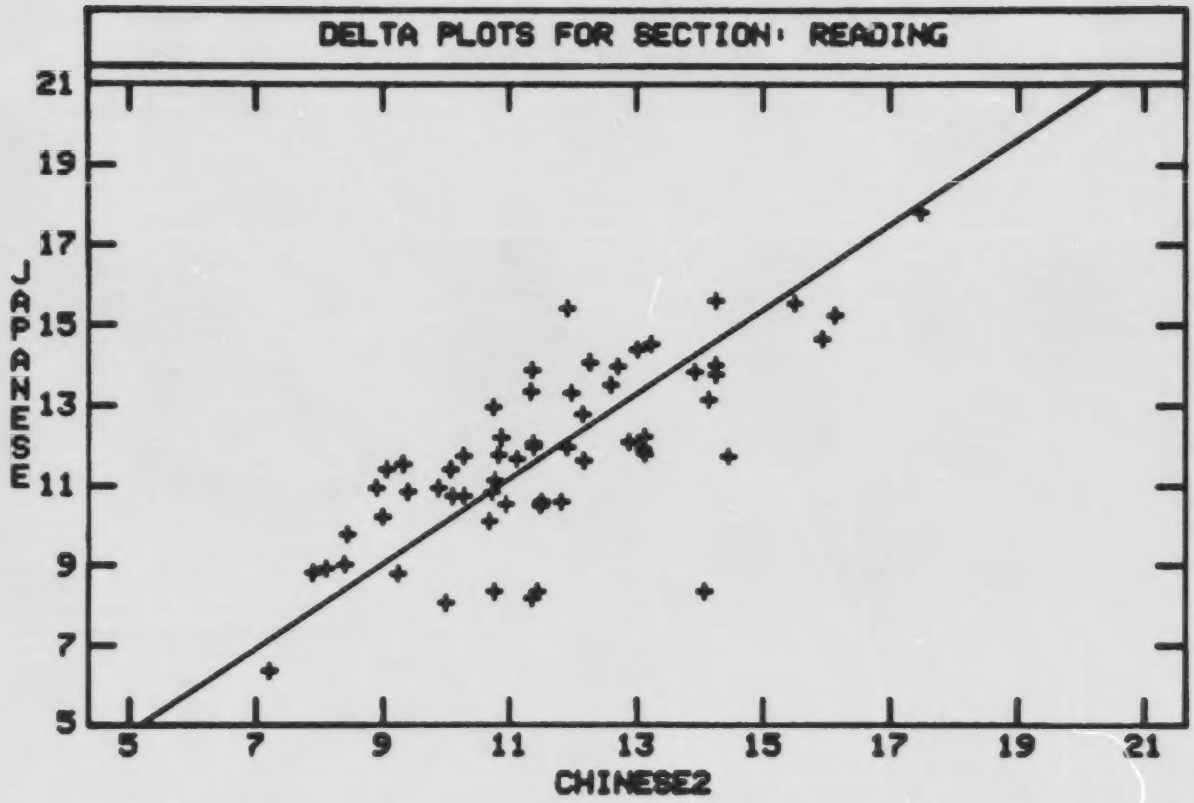
r = .77

CHINESE 1 & JAPANESE ADMIN 1



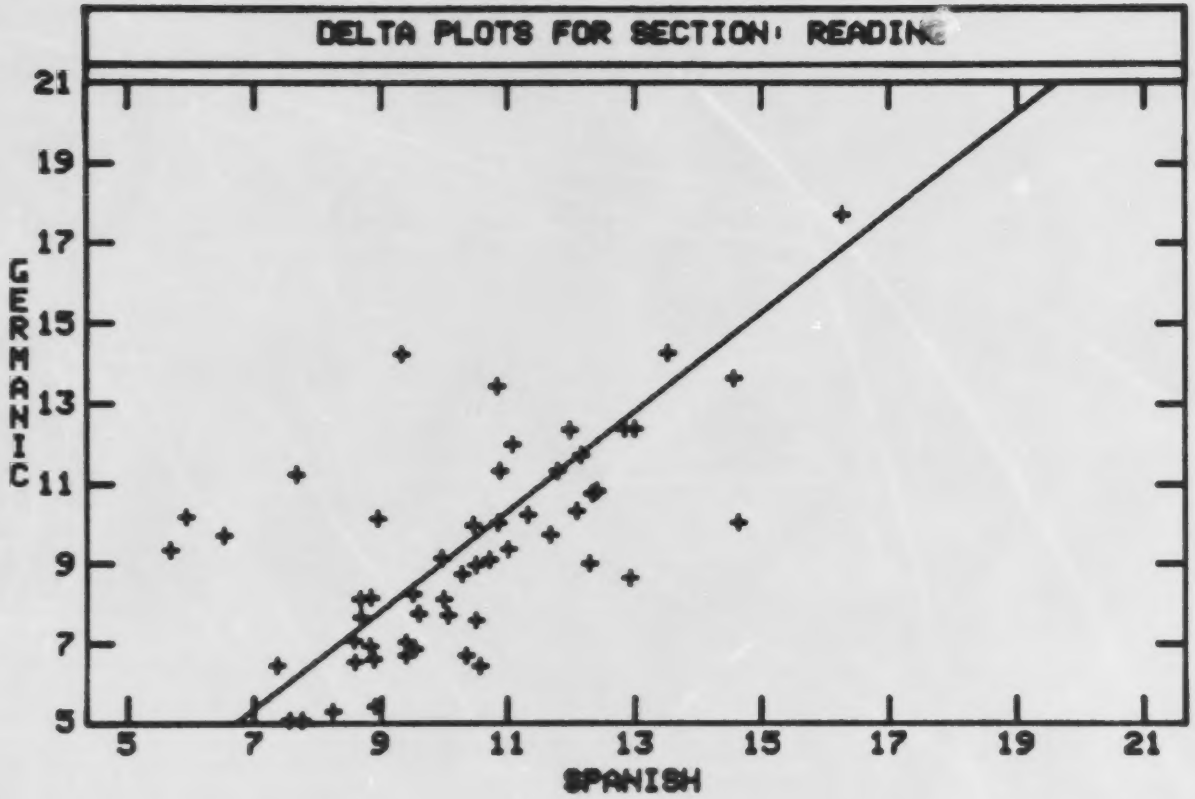
r = .73

CHINESE 2 & JAPANESE ADMIN 1



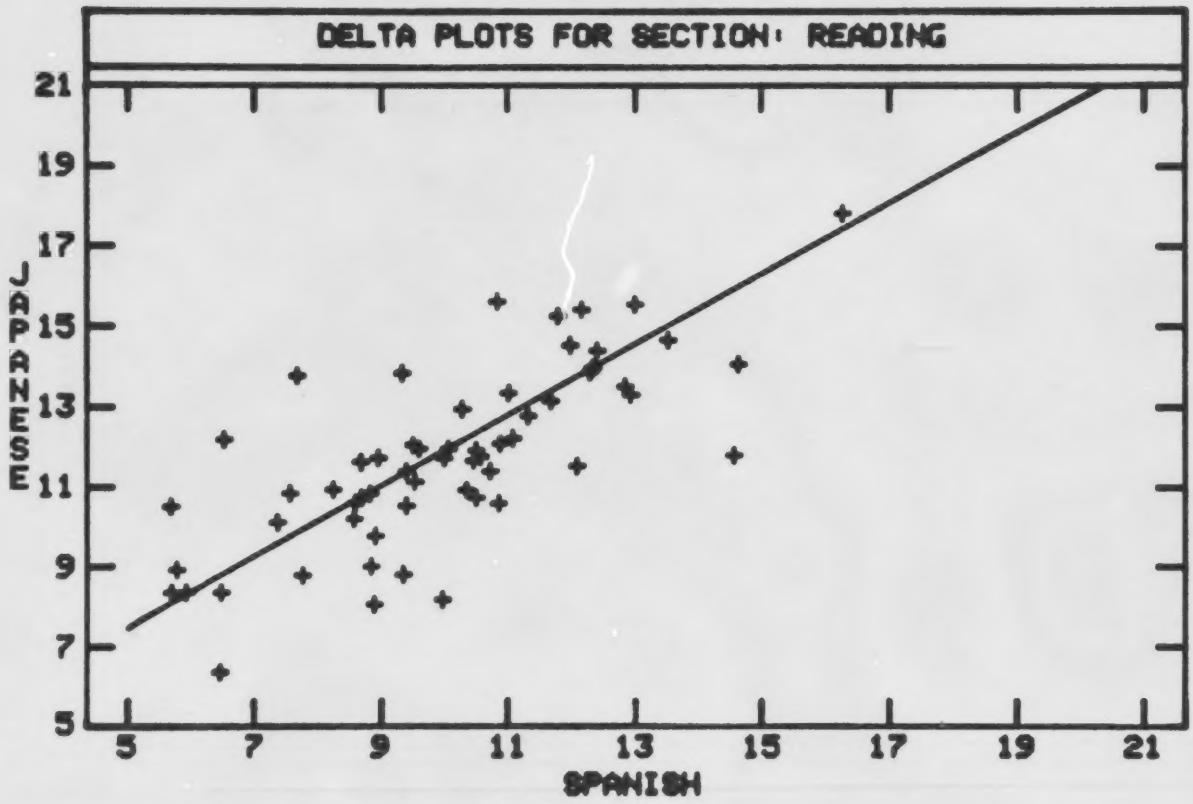
r = .73

SPANISH & GERMANIC ADMIN 1



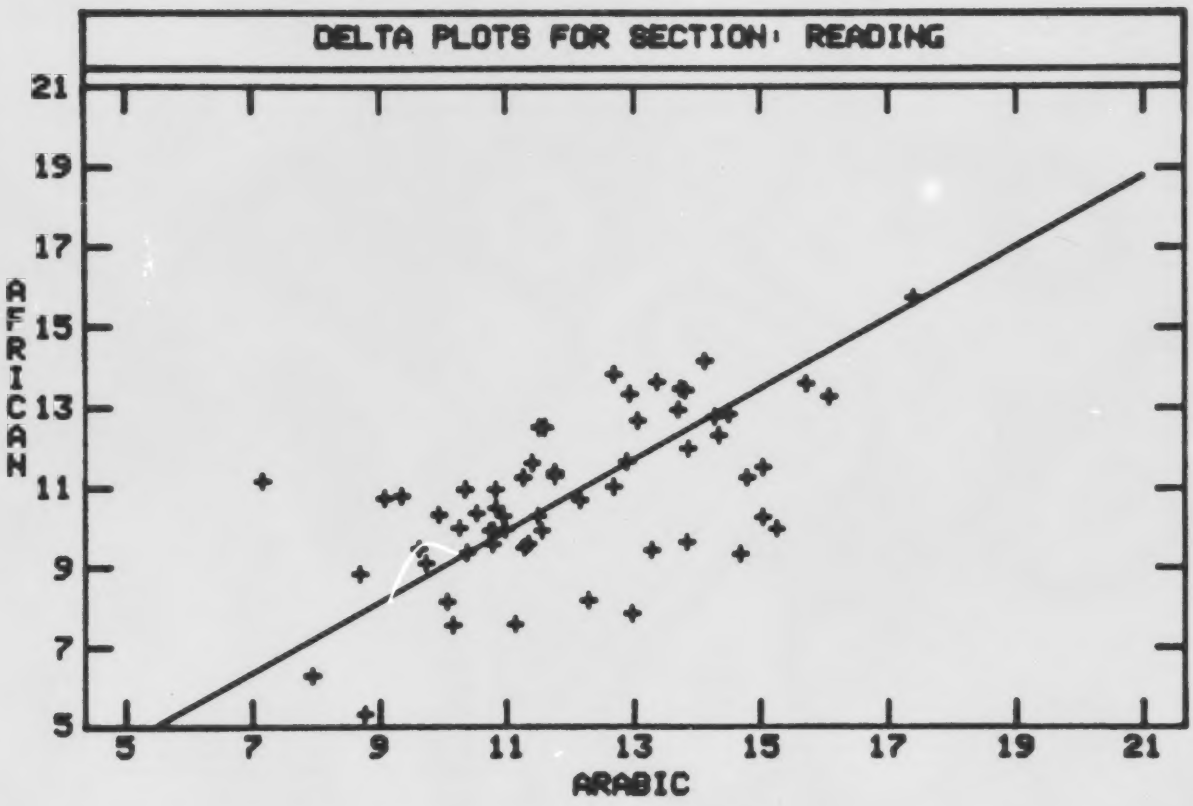
r = .67

SPANISH & JAPANESE ADMIN 1



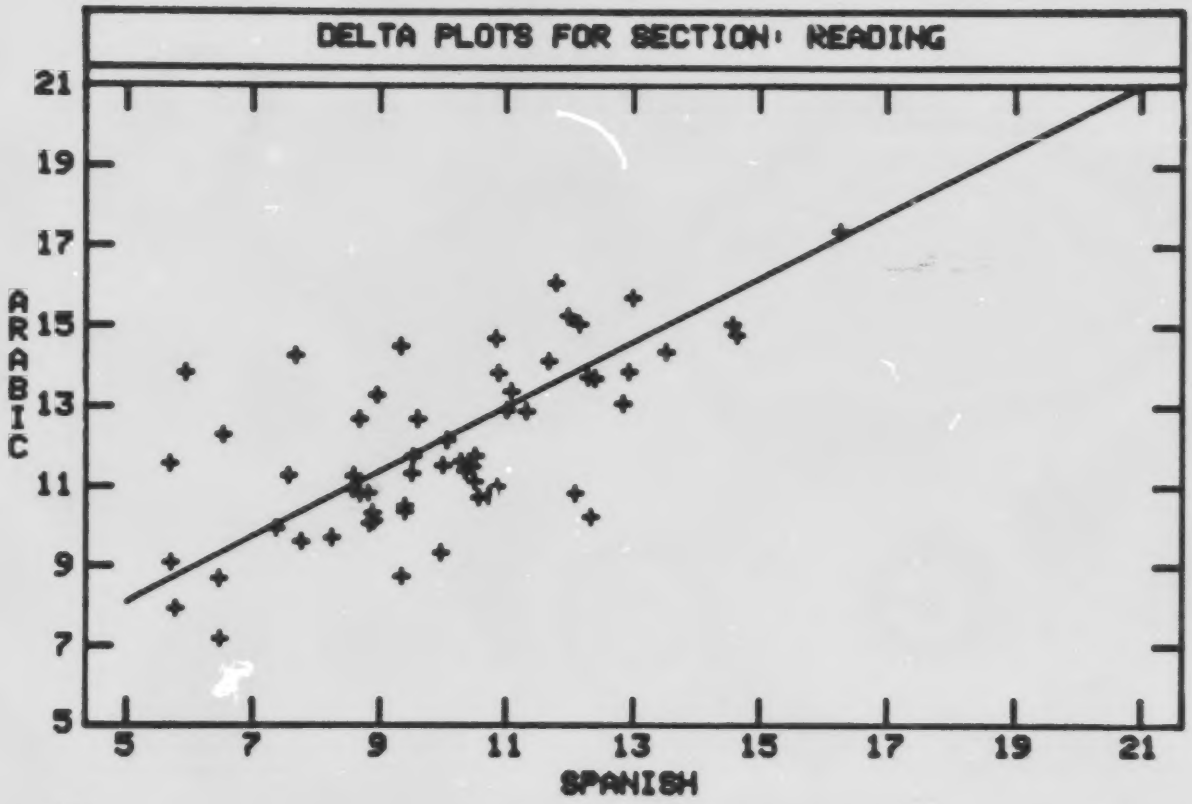
r = .65

ARABIC & AFRICAN ADMIN 1



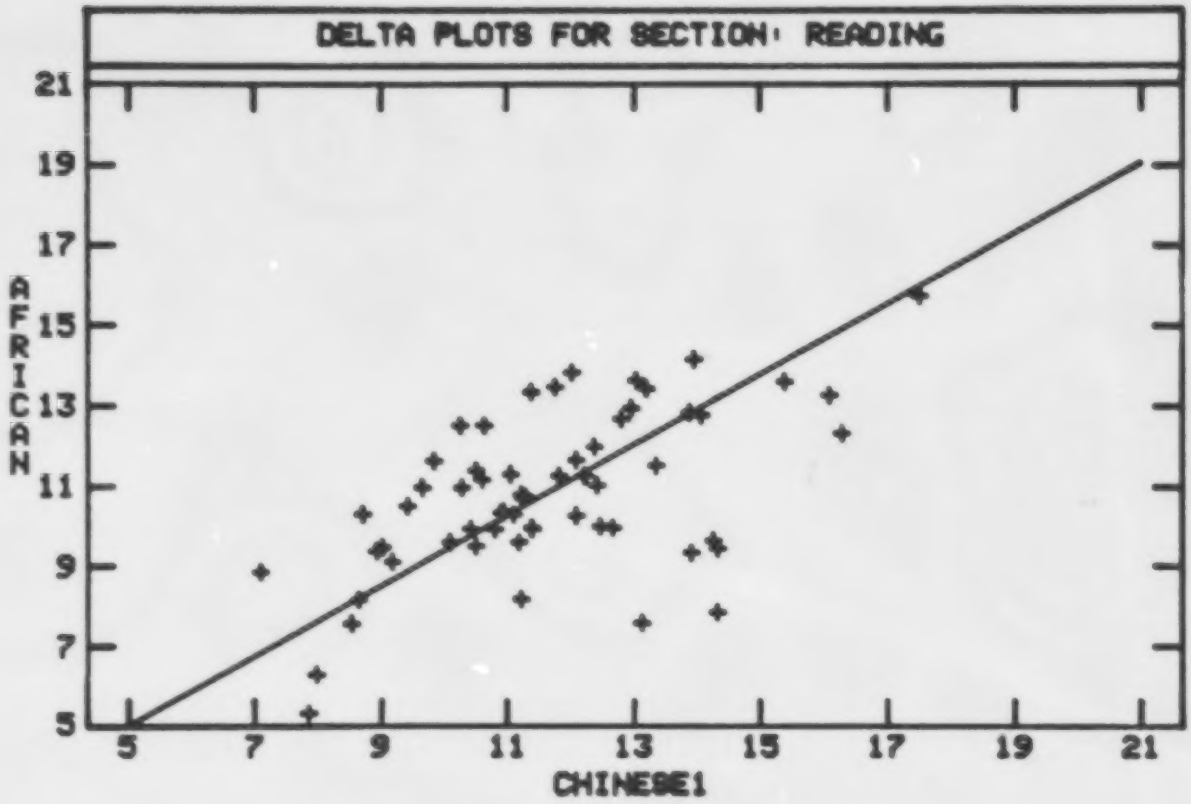
r = .61

SPANISH & ARABIC ADMIN 1



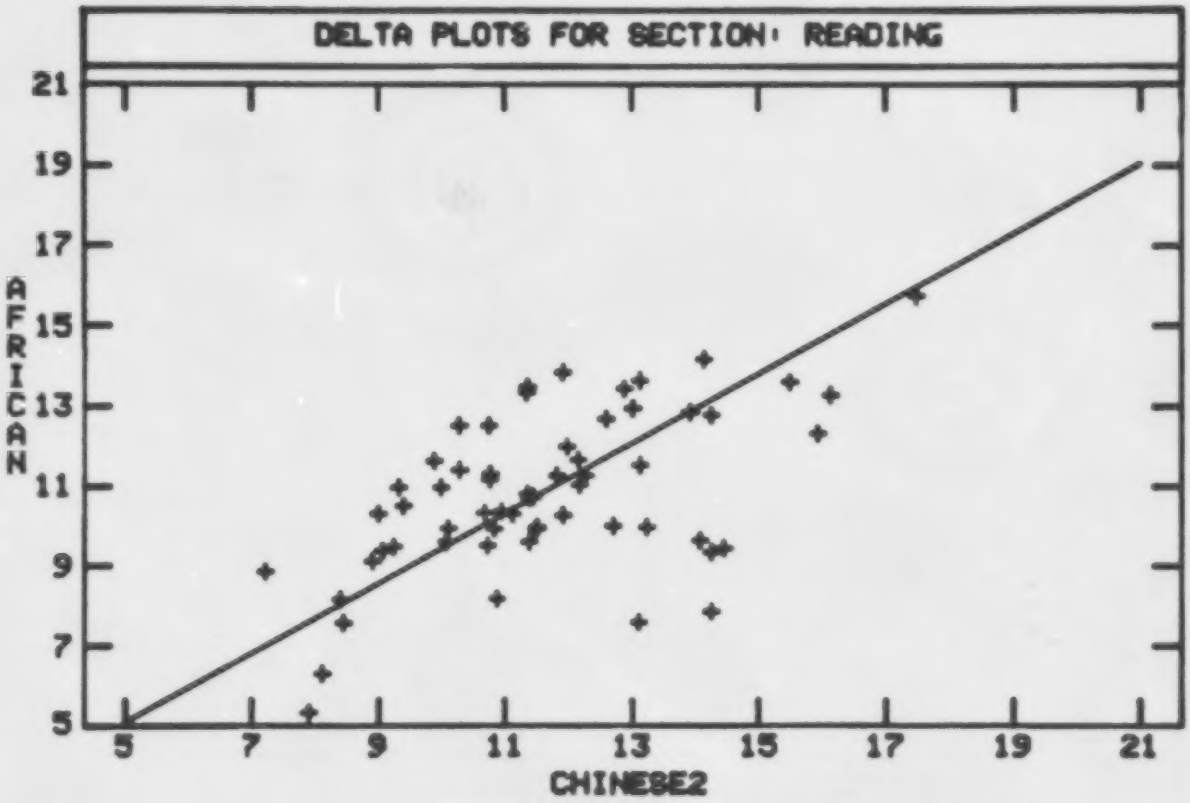
r = .60

CHINESE 1 & AFRICAN ADMIN 1



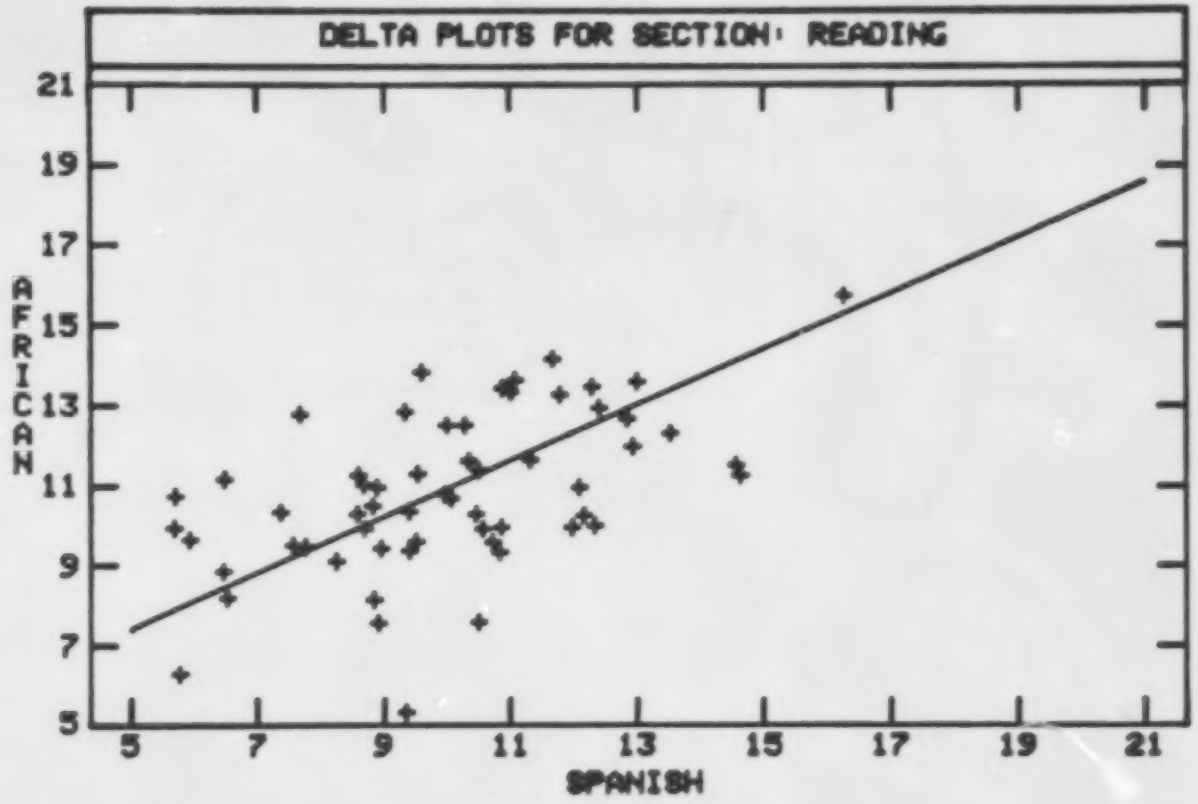
r = .57

CHINESE 2 & AFRICAN ADMIN 1



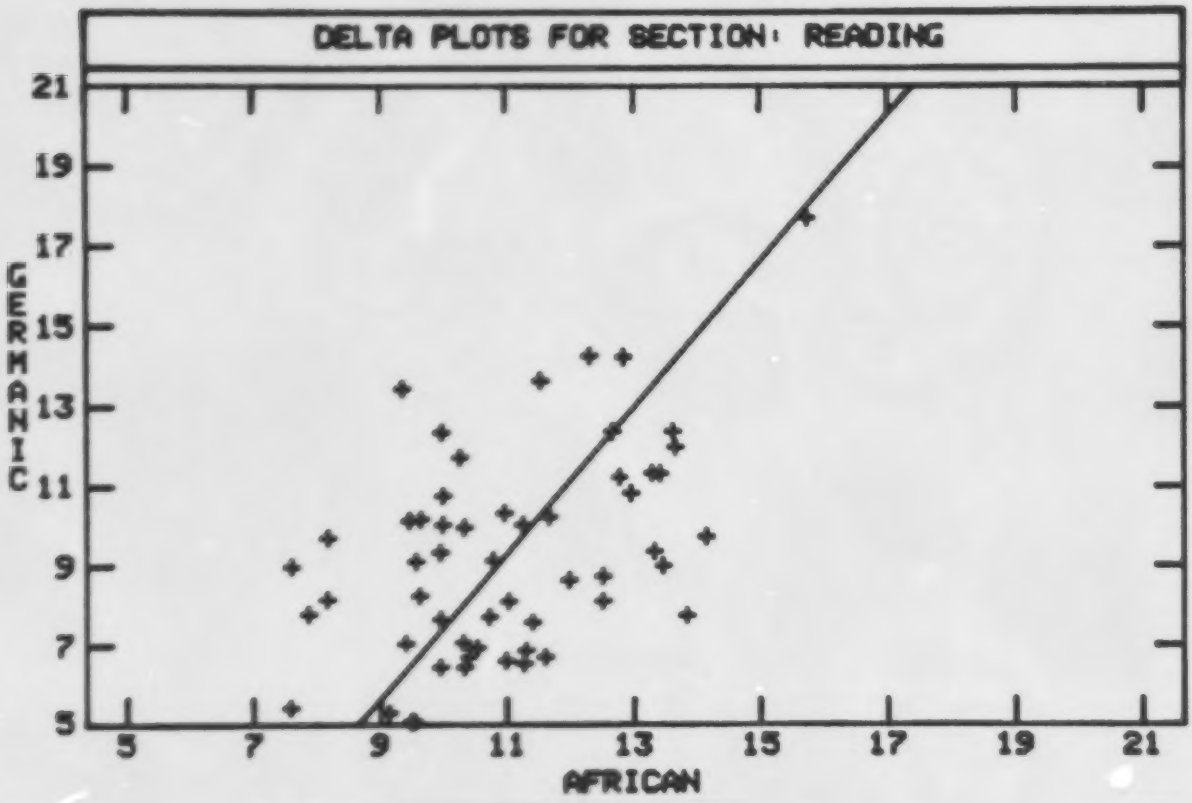
r = .56

SPANISH & ARFICAN ADMIN 1



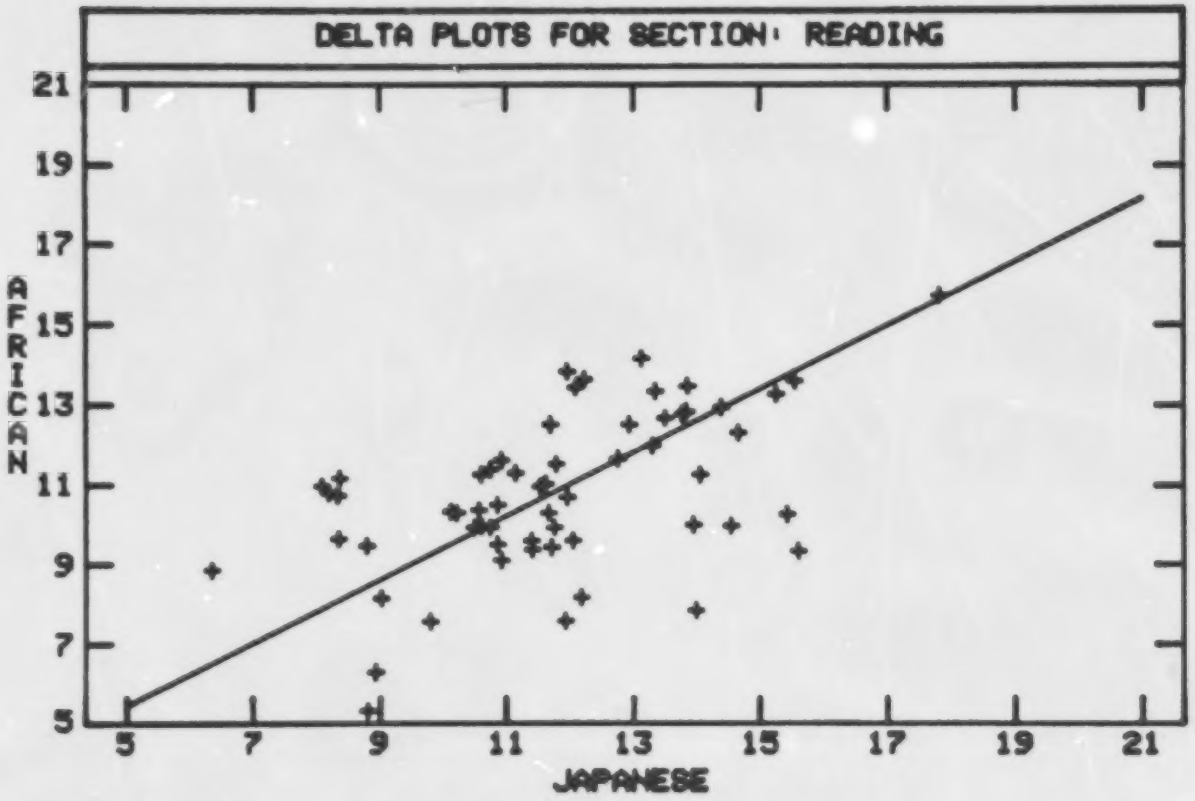
r = .55

AFRICAN & GERMANIC ADMIN 1



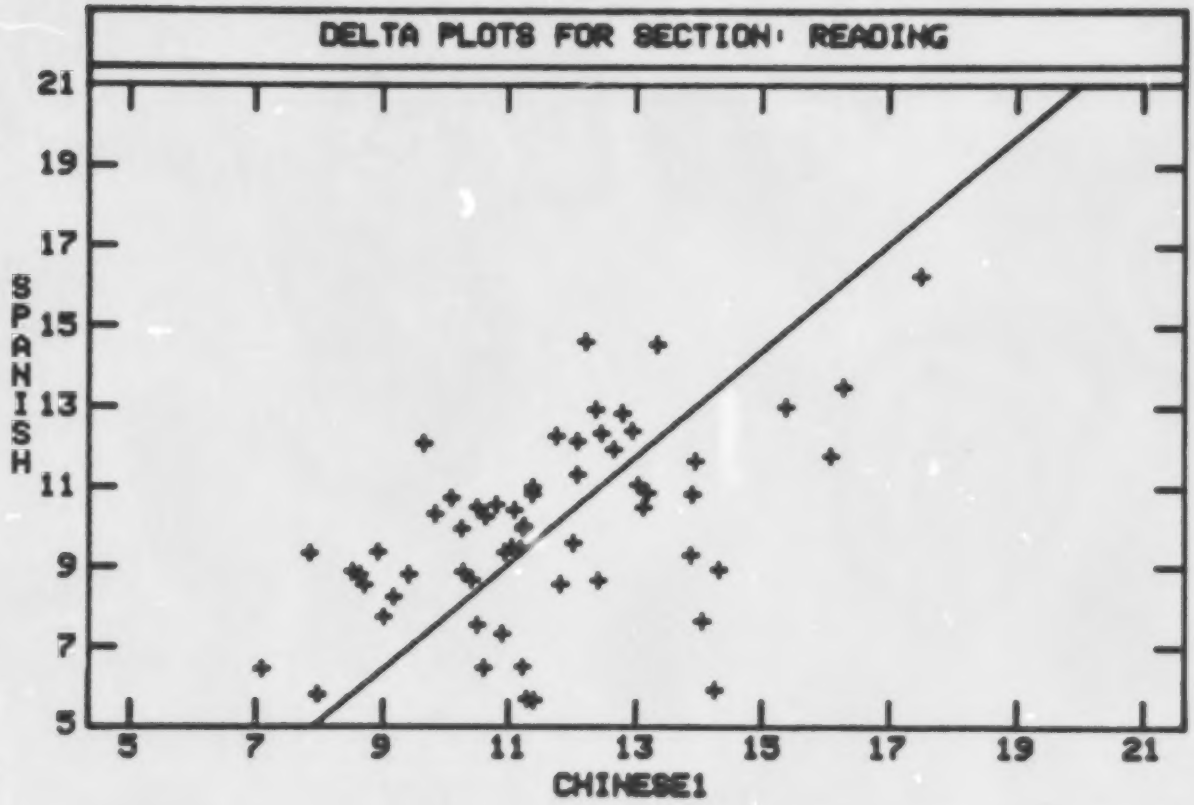
r = .54

JAPANESE & AFRICAN ADMIN 1



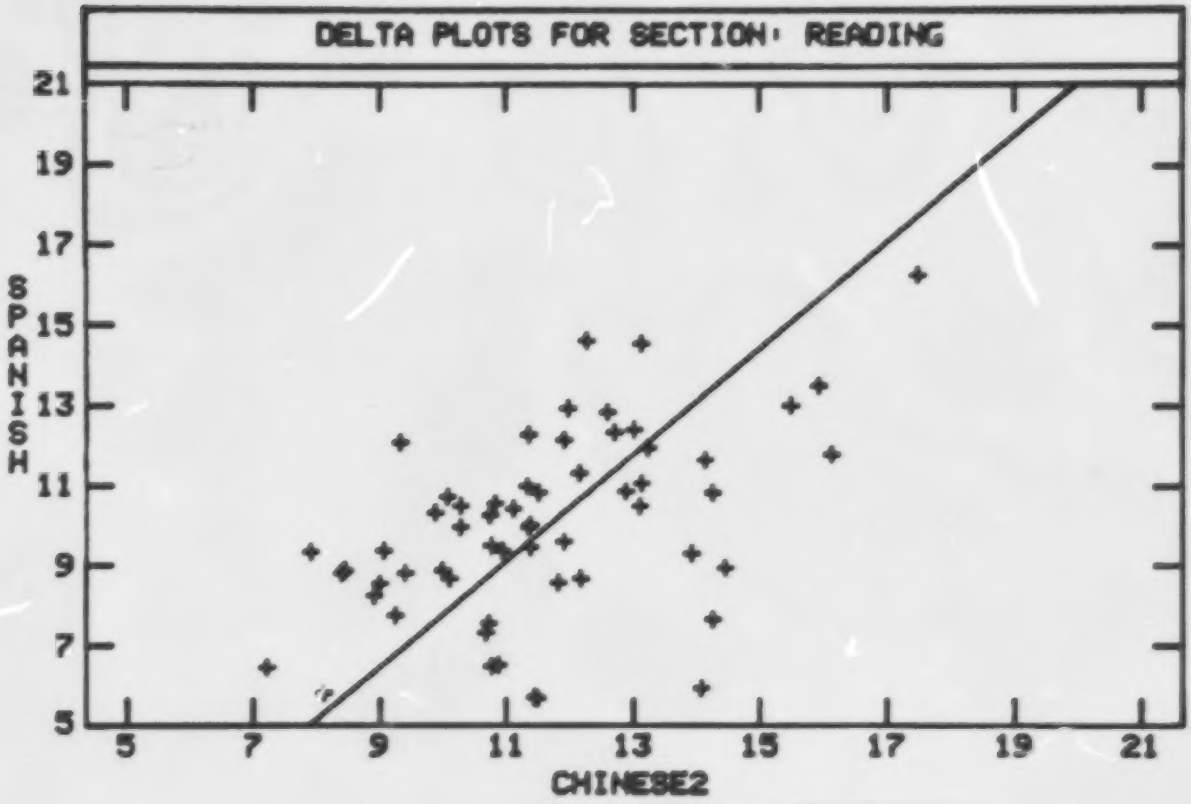
r = .52

CHINESE 1 & SPANISH ADMIN 1



r = .44

CHINESE 2 & SPANISH ADMIN 1



r = .43

Appendix C

Score Distributions by Test Section and Language Group

Score Distribution by Language Groups
TOEFL Section I: Listening Comprehension
Administration I

Score Level	Total Sample N	Percent	Chinese 1 N	Percent	Chinese 2 N	Percent	Spanish N	Percent	Japanese N	Percent	Arabic N	Percent	African N	Percent	Germanic N	Percent
46-50	633	10.5	51	5.1	52	5.2	162	16.3	28	2.8	52	7.6	13	1.8	275	46.1
43-45	549	9.1	73	7.3	62	6.2	129	13.0	49	4.9	65	9.5	37	5.0	134	22.5
40-42	601	10.0	87	8.7	92	9.2	136	13.7	86	8.6	68	9.9	49	6.6	83	13.9
37-39	607	10.1	104	10.4	106	10.6	99	10.0	107	10.7	82	12.0	68	9.2	41	6.9
34-36	635	10.6	107	10.7	136	13.6	97	9.8	123	12.3	70	10.2	80	10.8	22	3.7
31-33	596	9.9	113	11.3	117	11.7	80	8.1	122	12.2	61	8.9	82	11.1	21	3.5
28-30	606	10.0	123	12.3	117	11.7	62	6.3	115	11.5	70	10.2	111	15.0	8	1.3
25-27	535	8.9	105	10.5	117	11.7	59	6.0	111	11.1	56	8.2	79	10.7	8	1.3
21-24	589	9.8	122	12.2	93	9.3	79	8.0	106	10.6	81	11.8	105	14.2	3	.5
0-20	655	10.9	113	11.3	106	10.3	88	8.9	150	15.0	81	11.8	116	15.7	1	.2

Score Distribution by Language Groups
TOEFL Section II: Structure and Written Expression
Administration I

Score Level	Total Sample		Chinese 1		Chinese 2		Spanish		Japanese		Arabic		African		Germanic	
	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
35-39	568	9.5	62	6.2	58	5.8	95	9.6	28	2.8	25	3.6	147	19.9	153	25.7
32-34	676	11.3	95	9.5	107	10.7	116	11.7	53	5.3	50	7.3	119	16.1	136	22.8
30-31	574	9.6	84	8.4	87	8.7	93	9.4	74	7.4	59	8.6	88	11.9	89	14.9
28-29	586	9.8	113	11.3	89	8.9	85	8.6	81	8.1	58	8.5	91	12.3	69	11.6
26-27	622	10.4	103	10.3	116	11.6	70	7.1	117	11.7	69	10.1	93	12.6	54	9.1
24-25	582	9.7	122	12.2	100	10.0	92	9.3	104	10.4	66	9.6	63	8.5	35	5.9
22-23	617	10.3	119	11.9	120	12.0	94	9.5	128	12.8	76	11.1	53	7.2	27	4.5
20-21	509	8.5	89	8.9	98	9.8	102	10.3	99	9.9	68	9.9	35	4.7	18	3.0
17-19	574	9.6	100	10.0	107	10.7	105	10.6	146	14.7	82	12.0	28	3.8	6	1.0
0-16	698	11.6	111	11.1	116	11.6	139	14.0	167	16.8	133	19.4	23	3.1	9	1.5

Score Distribution by Language Groups
TOEFL Section III: Reading Comprehension and Vocabulary
Administration I

Score Level	Total Sample N	Total Sample Percent	Chinese 1 N	Chinese 1 Percent	Chinese 2 N	Chinese 2 Percent	Spanish N	Spanish Percent	Japanese N	Japanese Percent	Arabic N	Arabic Percent	African N	African Percent	Germanic N	Germanic Percent
54-60	545	9.1	51	5.1	55	5.5	172	17.4	25	2.5	31	4.5	73	9.9	138	23.2
50-53	636	10.6	60	6.0	63	6.3	180	18.2	77	7.7	38	5.5	78	10.5	140	23.5
47-49	548	9.1	68	6.8	74	7.4	130	13.1	79	7.9	21	3.1	76	10.3	100	16.8
44-46	616	10.3	89	8.9	82	8.2	130	13.1	99	9.9	56	8.2	79	10.7	81	13.6
41-43	621	10.3	111	11.1	108	10.8	90	9.1	97	9.7	69	10.1	87	11.8	59	9.9
38-40	634	10.6	117	11.7	114	11.4	89	9.0	103	10.3	80	11.7	100	13.5	31	5.2
35-37	542	9.0	120	12.0	111	11.1	59	6.0	97	9.7	63	9.2	72	9.7	20	3.4
31-34	607	10.1	115	11.5	137	13.7	63	6.4	113	11.3	88	12.8	80	10.8	11	1.8
26-30	623	10.4	138	13.8	124	12.4	44	4.4	138	13.8	109	15.9	58	7.8	12	2.0
0-25	634	10.6	129	12.9	130	13.0	34	3.4	169	17.0	131	19.1	37	5.0	4	.7

Score Distribution by Language Groups
TOEFL Section I: Listening Comprehension
Administration II

Score Level	Total Sample N	Total Sample Percent	Chinese 1 N	Chinese 1 Percent	Chinese 2 N	Chinese 2 Percent	Spanish N	Spanish Percent	Japanese N	Japanese Percent	Arabic N	Arabic Percent	African N	African Percent	Germanic N	Germanic Percent
46-50	599	9.4	16	1.6	14	1.4	131	13.2	28	2.8	42	4.3	10	2.2	358	37.3
42-45	716	11.2	52	5.2	55	5.5	155	15.6	50	5.0	76	7.7	17	3.8	311	32.4
39-41	573	9.0	69	6.9	52	5.2	129	13.0	98	9.8	84	8.5	16	3.6	125	13.0
36-38	612	9.6	102	10.3	90	9.0	1_3	11.4	93	9.3	112	11.3	29	6.5	73	7.6
33-35	643	10.1	115	11.6	115	11.5	98	9.9	120	12.0	115	11.7	37	8.2	43	4.5
30-32	674	10.6	134	13.5	128	12.9	86	8.7	135	13.6	118	12.0	49	10.9	24	2.5
27-29	647	10.2	139	14.0	150	15.1	55	5.5	124	12.4	103	10.4	59	13.1	17	1.8
24-26	651	10.2	125	12.6	133	13.4	79	8.0	124	12.4	96	9.7	88	19.6	6	.6
20-23	660	10.4	137	13.8	146	14.7	78	7.9	128	12.9	90	9.1	79	17.6	2	.2
0-19	598	9.4	104	10.5	113	11.3	68	6.9	96	9.6	151	15.3	65	14.5	1	.1

Score Distributions by Language Groups
TOEFL Section II: Structure and Written Expression
Administration II

Score Level	Total Sample N	Total Sample Percent	Chinese 1 N	Chinese 1 Percent	Chinese 2 N	Chinese 2 Percent	Spanish N	Spanish Percent	Japanese N	Japanese Percent	Arabic N	Arabic Percent	African N	African Percent	Germanic N	Germanic Percent
37-40	668	10.5	30	3.0	26	2.6	139	14.0	27	2.7	41	4.2	57	12.7	348	36.2
34-36	738	11.6	65	6.5	81	8.1	120	12.1	94	9.4	61	6.2	73	16.3	244	25.4
32-33	594	9.3	83	8.4	67	6.7	84	8.5	98	9.8	63	6.4	69	15.4	130	13.5
30-31	643	10.1	114	11.5	98	9.8	96	9.7	100	10.0	82	8.3	63	14.0	90	9.4
28-29	664	10.4	137	13.8	125	12.6	86	8.7	129	13.0	71	7.2	57	12.7	59	6.1
26-27	652	10.2	131	13.2	151	15.2	90	9.1	132	13.3	75	7.6	40	8.9	33	3.4
24-25	612	9.6	128	12.9	128	12.9	86	8.7	115	11.5	101	10.2	34	7.6	20	2.1
22-23	563	8.8	119	12.0	123	12.3	67	6.8	98	9.8	114	11.6	21	4.7	21	2.2
19-21	555	8.7	98	9.9	106	10.6	83	8.4	100	10.0	142	14.4	16	3.6	10	1.0
0-18	684	10.7	88	8.9	91	9.1	141	14.2	103	10.3	237	24.0	19	4.2	5	.5

Score Distributions by Language Groups
TOEFL Section III: Reading Comprehension and Vocabulary
Administration II

Score Level	Total Sample		Chinese 1		Chinese 2		Spanish		Japanese		Arabic		African		Germanic	
	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
51-60	681	10.7	60	6.0	56	5.6	169	17.0	32	3.2	34	3.4	46	10.2	284	29.6
47-50	678	10.6	108	10.9	107	10.7	124	12.5	57	5.7	36	3.6	49	10.9	197	20.5
44-46	664	10.4	113	11.4	117	11.7	101	10.2	83	8.3	35	3.5	58	12.9	157	16.3
41-43	696	10.9	134	13.5	106	10.6	132	13.3	91	9.1	60	6.1	60	13.4	113	11.8
38-40	735	11.5	129	13.0	156	15.7	123	12.4	98	9.8	77	7.8	68	15.1	84	8.7
35-37	652	10.2	123	12.4	142	14.2	95	9.6	103	10.3	79	8.0	59	13.1	51	5.3
32-34	585	9.2	101	10.2	98	9.8	76	7.7	110	11.0	121	12.3	44	9.8	35	3.6
29-31	521	8.2	97	9.8	75	7.5	58	5.8	111	11.1	125	12.7	31	6.9	24	2.5
24-28	635	10.0	82	8.3	92	9.2	72	7.3	157	15.8	199	20.2	22	4.9	11	1.1
0-23	526	8.3	46	4.6	47	4.7	42	4.2	154	15.5	221	22.4	12	2.7	4	.4

Appendix D

Percentage Deviations from Expected Frequencies
of Correct Item Responses

Administration I (Form YTF4)

TOEFL Section I: Listening Comprehension

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African
1	36	95.88	Percentage Deviation	1.24	.65	-3.15	2.12	-0.61	-0.49
			Proportion Correct	.92	.92	.91	.92	.91	.89
			Delta	7.33	7.33	7.75	7.41	7.68	8.17
2	13	215.01	Percentage Deviation	9.33	5.32	-8.92	13.95	-9.49	-16.89
			Proportion Correct	.73	.71	.67	.74	.62	.52
			Delta	10.59	10.82	11.29	10.46	11.83	12.77
3	44	74.47	Percentage Deviation	-2.86	-0.81	0.51	-1.36	-4.17	10.50
			Proportion Correct	.70	.73	.80	.70	.71	.76
			Delta	10.88	10.60	9.69	10.96	10.81	10.20
4	11	229.26	Percentage Deviation	-1.44	-0.72	8.58	-19.79	10.00	7.18
			Proportion Correct	.72	.73	.86	.57	.82	.74
			Delta	10.71	10.58	8.63	12.32	9.37	10.40
5	29	113.29	Percentage Deviation	5.44	3.18	-2.58	1.49	-4.20	-6.28
			Proportion Correct	.88	.87	.85	.84	.81	.76
			Delta	8.24	8.50	8.77	9.10	9.50	10.17
6	3[illegible]	97.12	Percentage Deviation	6.08	5.7[illegible]	-3.93	1.94	-3.44	-9.99
			Proportion Correct	.67	.6[illegible]	.69	.62	.63	.53
			Delta	11.26	11.18	11.06	11.78	11.69	12.72
7	41	79.73	Percentage Deviation	1.82	3.38	-0.44	-8.48	-1.57	6.54
			Proportion Correct	.70	.72	.76	.61	.70	.69
			Delta	10.89	10.68	10.20	11.87	10.96	10.98
8	7	299.27	Percentage Deviation	-2.45	3.21	-12.51	23.46	-4.83	-9.34
			Proportion Correct	.70	.75	.69	.86	.70	.61
			Delta	10.91	10.31	11.02	8.65	10.92	11.85
9	10	231.62	Percentage Deviation	8.85	4.03	-12.76	0.31	-5.98	5.97
			Proportion Correct	.80	.77	.67	.72	.69	.75
			Delta	9.68	10.05	11.21	10.66	10.97	10.34
10	8	292.11	Percentage Deviation	-6.05	3.31	-0.18	-23.31	-0.66	37.88
			Proportion Correct	.48	.53	.62	.37	.54	.64
			Delta	13.19	12.67	11.79	14.29	12.62	11.55

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African
11	40	83.22	Percentage Deviation	-0.15	0.78	-0.65	-1.16	-10.48	12.80
			Proportion Correct	.64	.66	.74	.61	.60	.66
			Delta	11.58	11.38	10.46	11.89	12.03	11.31
12	49	61.37	Percentage Deviation	2.68	-1.54	0.26	1.72	6.04	-11.05
			Proportion Correct	.58	.56	.67	.54	.62	.45
			Delta	12.24	12.39	11.23	12.55	11.74	13.49
13	18	147.72	Percentage Deviation	2.62	1.90	-3.85	-8.54	3.93	7.21
			Proportion Correct	.85	.85	.84	.74	.87	.86
			Delta	8.86	8.82	8.96	10.39	8.46	8.76
14	35	96.26	Percentage Deviation	-9.24	-3.32	4.44	-4.51	7.17	9.53
			Proportion Correct	.52	.56	.72	.53	.65	.58
			Delta	12.78	12.38	10.70	12.74	11.47	12.24
15	21	136.02	Percentage Deviation	-1.75	2.93	6.41	4.23	2.25	-22.02
			Proportion Correct	.50	.54	.66	.51	.55	.36
			Delta	12.97	12.64	11.30	12.93	12.47	14.45
16	9	235.66	Percentage Deviation	-26.70	-24.97	-1.12	44.34	-12.93	33.07
			Proportion Correct	.21	.21	.37	.38	.27	.33
			Delta	16.27	16.21	14.35	14.25	15.51	14.75
17	37	90.44	Percentage Deviation	8.87	4.79	-7.48	-5.79	-1.98	2.91
			Proportion Correct	.61	.60	.61	.51	.58	.53
			Delta	11.34	11.97	11.84	12.89	12.22	12.69
18	25	123.23	Percentage Deviation	1.63	6.23	-14.55	-5.71	5.72	16.15
			Proportion Correct	.38	.40	.39	.34	.42	.40
			Delta	14.18	14.01	14.08	14.69	13.78	13.98
19	17	149.57	Percentage Deviation	-1.77	-2.03	-8.83	11.48	-16.79	21.22
			Proportion Correct	.52	.52	.55	.56	.45	.60
			Delta	12.84	12.82	12.53	12.36	13.48	12.03
20	39	84.17	Percentage Deviation	12.77	10.18	-4.11	-6.14	-15.22	-1.76
			Proportion Correct	.44	.44	.46	.35	.35	.35
			Delta	13.56	13.61	13.36	14.54	14.51	14.55
21	1	503.13	Percentage Deviation	-11.95	-6.81	11.07	25.65	10.02	-36.08
			Proportion Correct	.60	.64	.84	.83	.77	.41
			Delta	12.00	11.59	9.06	9.13	10.05	13.87

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African
22	50	38.24	Percentage Deviation	0.44	0.38	-1.58	0.48	-0.55	0.95
			Proportion Correct	.93	.93	.93	.92	.92	.91
			Delta	7.16	7.04	7.11	7.44	7.34	7.51
23	30	110.64	Percentage Deviation	1.24	3.03	-0.48	3.07	2.37	-11.78
			Proportion Correct	.86	.88	.89	.86	.88	.73
			Delta	8.63	8.18	8.09	8.61	8.29	10.59
24	5	312.97	Percentage Deviation	-17.43	-20.16	11.65	12.67	15.18	2.99
			Proportion Correct	.53	.52	.80	.70	.76	.62
			Delta	12.69	12.78	9.63	10.88	10.18	11.78
25	47	69.19	Percentage Deviation	-2.29	-2.00	3.75	2.89	-1.36	-2.20
			Proportion Correct	.76	.77	.87	.78	.78	.72
			Delta	10.21	10.06	8.50	9.93	9.93	10.65
26	31	110.10	Percentage Deviation	.90	1.97	-3.05	7.66	-2.79	-7.39
			Proportion Correct	.85	.86	.86	.89	.82	.75
			Delta	8.89	8.59	8.76	8.16	9.28	10.30
27	45	74.00	Percentage Deviation	0.19	-2.95	4.67	-2.95	4.19	-3.32
			Proportion Correct	.75	.74	.86	.71	.80	.68
			Delta	10.28	10.42	8.70	10.81	9.69	11.10
28	46	73.79	Percentage Deviation	-1.94	0.38	-0.99	-1.91	5.23	1.16
			Proportion Correct	.69	.72	.77	.67	.76	.68
			Delta	10.99	10.72	10.02	11.20	10.15	11.16
29	2	411.45	Percentage Deviation	14.71	17.37	2.20	-4.35	-1.60	-42.84
			Proportion Correct	.73	.76	.73	.58	.64	.34
			Delta	10.58	10.23	10.49	12.15	11.54	14.69
30	3	340.49	Percentage Deviation	9.12	9.79	-5.52	-28.75	15.80	5.17
			Proportion Correct	.66	.68	.67	.41	.73	.58
			Delta	11.33	11.15	11.25	13.87	10.53	12.16
31	27	116.23	Percentage Deviation	2.69	-1.81	12.04	-4.36	0.86	-19.50
			Proportion Correct	.49	.47	.68	.43	.51	.33
			Delta	13.14	13.26	11.15	13.75	12.87	14.72
32	20	136.59	Percentage Deviation	-11.98	-5.50	3.91	-13.28	14.89	21.49
			Proportion Correct	.35	.38	.54	.32	.49	.41
			Delta	14.59	14.23	12.56	14.91	13.09	13.94

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African
33	4	327.56	Percentage Deviation	-10.94	-12.00	-7.87	-8.88	16.59	42.17
			Proportion Correct	.52	.51	.63	.51	.71	.76
			Delta	12.84	12.86	11.69	12.94	10.81	10.20
34	32	108.48	Percentage Deviation	7.66	3.45	1.56	-11.21	2.98	-6.34
			Proportion Correct	.65	.64	.72	.52	.65	.52
			Delta	11.45	11.62	10.71	12.83	11.47	12.78
35	15	169.82	Percentage Deviation	-5.05	-7.39	-13.58	-3.18	14.84	30.97
			Proportion Correct	.46	.45	.50	.45	.58	.58
			Delta	13.41	13.46	13.04	13.54	12.19	12.22
36	24	126.04	Percentage Deviation	-7.62	-6.98	6.87	-3.45	8.76	5.86
			Proportion Correct	.61	.62	.80	.62	.74	.65
			Delta	11.88	11.76	9.62	11.82	10.39	11.45
37	26	117.31	Percentage Deviation	-1.41	4.23	-3.74	7.94	-18.17	8.34
			Proportion Correct	.61	.64	.63	.65	.51	.65
			Delta	11.92	11.55	11.66	11.43	12.88	11.46
38	23	130.05	Percentage Deviation	-0.77	0.55	4.63	2.26	-10.38	0.50
			Proportion Correct	.85	.87	.93	.86	.77	.83
			Delta	8.89	8.56	6.96	8.70	10.03	9.17
39	48	68.14	Percentage Deviation	10.19	7.51	4.69	-2.16	-16.96	-14.75
			Proportion Correct	.28	.27	.36	.22	.23	.18
			Delta	15.38	15.46	14.40	16.04	15.97	16.62
40	16	158.57	Percentage Deviation	3.15	0.51	4.42	4.70	-10.03	-8.35
			Proportion Correct	.84	.83	.91	.84	.74	.72
			Delta	8.96	9.17	7.71	9.03	10.37	10.67
41	33	106.10	Percentage Deviation	0.97	3.42	6.91	4.26	-13.78	-9.33
			Proportion Correct	.65	.68	.77	.65	.57	.55
			Delta	11.42	11.15	10.02	11.43	12.27	12.50
42	6	307.09	Percentage Deviation	-10.25	-16.77	5.37	24.86	-0.66	-3.56
			Proportion Correct	.62	.58	.81	.85	.71	.64
			Delta	11.75	12.18	9.54	8.93	10.81	11.58
43	14	180.25	Percentage Deviation	-8.26	-12.90	8.71	20.39	-4.81	-7.54
			Proportion Correct	.49	.47	.71	.62	.54	.45
			Delta	13.08	13.25	10.83	11.82	12.61	13.54

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African
44	19	140.97	Percentage Deviation	-4.60	-3.13	9.09	-4.93	11.44	-8.79
			Proportion Correct	.63	.65	.82	.60	.76	.55
			Delta	11.71	11.50	9.29	11.97	10.18	12.47
45	28	115.03	Percentage Deviation	0.81	-0.44	-7.61	2.28	0.33	7.27
			Proportion Correct	.82	.82	.80	.81	.83	.84
			Delta	9.33	9.37	9.70	9.44	9.26	9.02
46	38	87.19	Percentage Deviation	1.75	1.71	1.35	-7.94	-0.15	3.60
			Proportion Correct	.37	.37	.47	.31	.38	.33
			Delta	14.37	14.32	13.35	14.97	14.17	14.76
47	22	133.50	Percentage Deviation	12.72	8.55	-3.24	4.00	-8.19	-22.80
			Proportion Correct	.58	.56	.61	.51	.50	.36
			Delta	12.19	12.36	11.91	12.92	12.99	14.45
48	42	79.53	Percentage Deviation	2.84	7.68	3.50	-11.30	1.75	-9.61
			Proportion Correct	.42	.45	.57	.34	.45	.32
			Delta	13.80	13.54	12.34	14.69	13.45	14.90
49	43	77.37	Percentage Deviation	2.84	-0.25	-1.03	-9.65	1.89	9.63
			Proportion Correct	.66	.66	.74	.56	.69	.65
			Delta	11.32	11.40	10.42	12.38	11.08	11.47
50	12	215.74	Percentage Deviation	2.48	3.47	4.49	-17.90	8.71	18.02
			Proportion Correct	.70	.72	.81	.55	.65	.76
			Delta	10.88	10.69	9.50	12.53	11.50	10.20

Administration I (Form YTF4)

TOEFL Section II: Structure and Written Expression

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
1	31	158.88	Percentage Deviation	-7.31	-10.28	4.19	5.46	5.51	5.21	0.47
			Proportion Correct	.74	.71	.83	.79	.80	.93	.92
			Delta	10.47	10.79	9.24	9.74	9.63	7.22	7.35
2	33	127.58	Percentage Deviation	-5.25	-6.01	-2.63	6.84	2.74	7.22	-0.46
			Proportion Correct	.70	.69	.72	.75	.73	.87	.84
			Delta	10.88	10.98	10.68	10.29	10.58	8.44	9.01
3	37	99.24	Percentage Deviation	-0.45	-0.09	-1.95	-0.41	5.12	-1.71	1.22
			Proportion Correct	.91	.91	.89	.88	.93	.94	.98
			Delta	7.71	7.71	8.18	8.31	7.09	6.76	4.33
4	24	208.56	Percentage Deviation	-8.03	-8.72	3.20	2.11	6.78	9.39	-0.82
			Proportion Correct	.73	.72	.82	.77	.81	.96	.90
			Delta	10.51	10.63	9.38	10.03	9.46	5.96	7.77
5	30	161.14	Percentage Deviation	0.45	0.46	3.84	1.48	-3.56	-8.28	4.36
			Proportion Correct	.84	.84	.86	.81	.77	.84	.98
			Delta	8.99	9.06	8.61	9.51	10.01	9.08	4.50
6	23	213.80	Percentage Deviation	-5.80	-4.69	17.86	-9.61	4.15	0.57	-2.17
			Proportion Correct	.64	.65	.31	.58	.67	.77	.79
			Delta	11.56	11.51	9.54	12.21	11.19	10.08	9.84
7	22	226.41	Percentage Deviation	-5.99	-4.92	4.68	-2.68	12.33	18.39	1.39
			Proportion Correct	.66	.66	.74	.63	.58	.95	.86
			Delta	11.36	11.33	10.48	11.65	12.22	6.31	8.60
8	1	795.30	Percentage Deviation	37.78	35.15	-35.00	-59.62	7.42	19.16	-7.79
			Proportion Correct	.64	.63	.31	.17	.46	.70	.59
			Delta	11.52	11.69	14.96	16.84	13.39	10.92	12.05
9	20	227.81	Percentage Deviation	9.08	6.70	-16.01	-3.53	-7.88	18.67	-9.71
			Proportion Correct	.52	.51	.42	.43	.42	.68	.57
			Delta	12.77	12.90	13.84	13.74	13.84	11.09	12.34
10	2	601.09	Percentage Deviation	-26.35	-25.44	18.59	-16.28	16.01	26.90	11.33
			Proportion Correct	.38	.38	.63	.37	.53	.87	.86
			Delta	14.22	14.20	11.71	14.36	12.69	8.44	8.67

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
11	21	227.75	Percentage Deviation	-6.51	-6.23	22.68	-11.80	0.59	-5.11	6.20
			Proportion Correct	.55	.55	.73	.47	.54	.68	.82
			Delta	12.49	12.53	10.60	13.35	12.58	11.15	9.33
12	5	571.67	Percentage Deviation	-26.53	-20.77	2.17	-18.69	-0.15	48.32	12.51
			Proportion Correct	.31	.33	.40	.29	.37	.85	.73
			Delta	15.01	14.77	13.56	15.26	14.35	8.85	10.53
14	8	490.34	Percentage Deviation	49.10	59.25	-40 91	-24.55	-29.52	1.50	-29.56
			Proportion Correct	.42	.44	.17	.18	.17	.40	.32
			Delta	13.85	13.61	16.78	16.72	16.78	14.01	14.92
15	11	343.01	Percentage Deviation	-13.80	-10.59	10.15	15.31	10.11	21.55	-30.49
			Proportion Correct	.39	.40	.51	.46	.45	.72	.46
			Delta	14.09	13.96	12.87	13.44	13.50	10.68	14.44
16	38	83.20	Percentage Deviation	0.82	0.17	0.55	-4.57	3.69	0.25	0.40
			Proportion Correct	.94	.93	.93	.87	.94	.97	.99
			Delta	6.82	7.10	7.11	8.57	6.66	5.29	3.94
17	25	205.37	Percentage Deviation	1.92	3.83	-9.40	0.87	4.73	-1.95	1.64
			Proportion Correct	.92	.93	.81	.88	.91	.93	.99
			Delta	7.49	7.13	9.53	8.41	7.61	6.98	3.94
18	27	184.64	Percentage Deviation	4.14	4.56	5.23	-0.90	-13.14	-7.50	2.55
			Proportion Correct	.82	.82	.83	.75	.66	.80	.92
			Delta	9.28	9.28	9.17	10.38	11.39	9.60	7.25
19	18	247.77	Percentage Deviation	-10.20	-6.92	-5.52	24.57	-3.61	-6.24	10.54
			Proportion Correct	.59	.60	.62	.74	.58	.72	.91
			Delta	12.13	11.97	11.81	10.42	12.18	10.67	7.69
20	9	367.13	Percentage Deviation	-2.89	-3.65	9.29	9.65	2.86	-22.37	5.87
			Proportion Correct	.76	.76	.8[illegible]	.83	.79	.65	.92
			Delta	10.13	10.23	8.67	9.16	9.85	11.42	7.44
21	34	124.57	Percentage Deviation	1.13	3.12	-6.63	5.10	1.75	-4.72	0.33
			Proportion Correct	.81	.82	.74	.79	.77	.84	.92
			Delta	9.55	9.39	10.46	9.80	10.06	9.06	7.48
22	16	270.92	Percentage Deviation	6.15	6.06	-14.67	3.03	2.06	0.37	-3.36
			Proportion Correct	.89	.88	.70	.81	.81	.92	.92
			Delta	8.18	8.32	10.88	9.51	9.55	7.30	7.39

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
23	35	123.04	Percentage Deviation	-1.66	-3.44	1.74	-2.50	13.24	-7.26	4.53
			Proportion Correct	.68	.66	.71	.62	.74	.74	.89
			Delta	11.15	11.33	10.84	11.76	10.49	10.43	8.18
24	10	347.73	Percentage Deviation	9.98	14.34	8.56	-4.01	-9.06	-9.92	-20.07
			Proportion Correct	.73	.75	.72	.59	.57	.69	.65
			Delta	10.60	10.31	10.71	12.16	12.41	11.04	11.50
25	3	584.79	Percentage Deviation	12.68	12.99	-17.70	19.31	-4.74	-23.86	-4.31
			Proportion Correct	.81	.81	.59	.80	.65	.62	.82
			Delta	9.48	9.54	12.11	9.64	11.53	11.81	9.43
26	12	321.38	Percentage Deviation	-7.50	-7.10	8.68	25.46	-0.95	-24.98	8.03
			Proportion Correct	.52	.51	.62	.63	.52	.52	.81
			Delta	12.84	12.86	11.82	11.72	12.98	12.85	9.46
27	17	253.60	Percentage Deviation	9.64	9.01	-9.74	-9.40	4.01	-5.10	1.99
			Proportion Correct	.86	.85	.71	.66	.78	.84	.95
			Delta	8.67	8.86	10.89	11.34	10.06	8.99	6.58
28	36	109.87	Percentage Deviation	-3.18	-3.13	2.45	0.11	1.05	1.78	2.38
			Proportion Correct	.77	.76	.82	.75	.78	.89	.93
			Delta	10.10	10.14	9.45	10.28	10.06	8.13	7.13
29	39	82.58	Percentage Deviation	1.16	-1.15	-4.80	-1.39	4.11	3.32	0.75
			Proportion Correct	.78	.76	.74	.71	.77	.92	.94
			Delta	9.88	10.18	10.53	10.84	10.20	7.52	6.93
30	7	509.99	Percentage Deviation	4.73	2.88	-24.38	14.18	7.63	26.20	-30.37
			Proportion Correct	.56	.54	.41	.55	.54	.84	.50
			Delta	12.44	12.59	14.00	12.61	12.82	9.15	13.01
31	32	151.23	Percentage Deviation	5.02	5.10	3.54	0.83	-2.12	-20.58	6.80
			Proportion Correct	.58	.58	.59	.51	.52	.53	.76
			Delta	12.17	12.21	12.22	12.94	13.00	12.75	10.14
32	15	295.12	Percentage Deviation	-23.66	-22.75	42.66	3.16	-7.36	-2.29	8.33
			Proportion Correct	.29	.29	.57	.36	.34	.47	.56
			Delta	15.20	15.20	12.45	14.51	14.87	13.46	12.46
33	28	176.78	Percentage Deviation	7.73	8.66	-1.55	-6.63	-7.26	-5.19	0.54
			Proportion Correct	.82	.83	.77	.67	.69	.84	.93
			Delta	9.27	9.26	10.27	11.31	11.24	9.17	7.26

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
34	19	243.93	Percentage Deviation	20.14	14.51	-8.19	0.91	-19.39	-21.11	5.42
			Proportion Correct	.64	.61	.51	.50	.42	.52	.74
			Delta	11.57	11.93	13.05	13.13	14.07	12.96	10.44
35	6	524.43	Percentage Deviation	11.34	12.17	-16.49	28.33	-20.73	-20.90	-2.12
			Proportion Correct	.74	.74	.58	.80	.52	.62	.82
			Delta	10.44	10.43	12.43	9.72	13.09	11.89	9.47
36	4	572.36	Percentage Deviation	1.12	1.66	22.27	-54.12	-11.92	14.00	17.50
			Proportion Correct	.49	.49	.63	.20	.41	.73	.84
			Delta	13.08	13.10	11.86	16.52	14.22	10.62	9.15
37	14	298.28	Percentage Deviation	-15.64	-21.40	12.47	11.75	17.72	18.21	23.59
			Proportion Correct	.35	.32	.50	.40	.45	.47	.80
			Delta	14.59	14.89	13.23	14.20	13.76	13.43	9.69
38	29	169.53	Percentage Deviation	11.84	4.17	-16.96	22.90	-20.94	-11.03	5.95
			Proportion Correct	.46	.43	.37	.46	.32	.48	.64
			Delta	13.40	13.74	14.54	13.53	15.18	13.29	11.66
39	13	306.22	Percentage Deviation	-16.77	-23.57	31.78	-30.91	-2.06	26.63	7.60
			Proportion Correct	.27	.25	.49	.20	.31	.58	.57
			Delta	15.40	15.71	13.38	16.48	15.25	12.28	12.44
40	26	195.67	Percentage Deviation	-2.48	3.28	-19.00	7.76	-2.73	31.93	-19.62
			Proportion Correct	.31	.32	.29	.29	.30	.61	.42
			Delta	14.98	14.84	15.47	15.28	15.45	12.05	13.91

Administration I (Form YTF4)

Section III: Reading Comprehension & Vocabulary

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
1	14	499.46	Percentage Deviation	-13.48	-15.50	10.53	14.95	12.21	-13.30	6.67
			Proportion Correct	.67	.65	.97	.88	.84	.71	.98
			Delta	11.29	11.45	5.71	8.37	9.10	10.75	4.93
2	22	361.68	Percentage Deviation	3.36	2.86	-12.10	-1.18	0.38	7.21	3.02
			Proportion Correct	.90	.90	.82	.85	.85	.97	.98
			Delta	7.88	7.92	9.36	8.83	8.78	5.29	4.65
3	37	220.17	Percentage Deviation	7.78	7.51	-12.28	-6.26	4.37	4.55	-2.40
			Proportion Correct	.77	.77	.71	.65	.71	.80	.83
			Delta	10.10	10.09	10.74	11.42	10.79	9.58	9.12
4	35	253.00	Percentage Deviation	-8.69	-9.13	6.10	9.75	5.08	-10.86	8.58
			Proportion Correct	.61	.61	.86	.72	.67	.67	.95
			Delta	11.84	11.84	8.61	10.62	11.27	11.27	6.56
5	44	174.47	Percentage Deviation	-10.09	-6.72	11.53	4.43	-8.86	-1.40	6.38
			Proportion Correct	.56	.58	.86	.63	.53	.69	.89
			Delta	12.44	12.19	8.69	11.63	12.69	11.04	8.14
6	13	511.77	Percentage Deviation	-9.15	-11.05	7.18	10.36	19.09	-19.63	6.71
			Proportion Correct	.73	.71	.95	.88	.93	.68	.98
			Delta	10.60	10.76	6.50	8.39	7.18	11.16	4.65
7	25	338.10	Percentage Deviation	-8.71	-10.64	-3.73	20.92	13.22	-7.55	-0.68
			Proportion Correct	.67	.66	.78	.88	.82	.71	.83
			Delta	11.23	11.37	9.97	8.21	9.35	10.79	9.15
8	26	314.32	Percentage Deviation	-7.41	-8.78	16.59	4.96	-5.71	1.01	-6.77
			Proportion Correct	.65	.65	.97	.73	.64	.78	.82
			Delta	11.41	11.49	5.68	10.51	11.57	9.95	9.36
9	42	183.08	Percentage Deviation	3.62	3.07	-11.55	-7.70	10.18	2.63	7.67
			Proportion Correct	.71	.71	.73	.62	.71	.78	.95
			Delta	10.82	10.83	10.58	11.77	10.74	9.95	6.43
10	15	481.44	Percentage Deviation	-20.81	-20.69	-0.91	-0.21	15.54	34.41	5.92
			Proportion Correct	.49	.49	.73	.60	.68	.91	.84
			Delta	13.13	13.10	10.52	11.94	11.14	7.62	9.01

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
11	50	134.68	Percentage Deviation	-2.15	-4.41	-8.03	9.05	6.21	7.70	-4.90
			Proportion Correct	.66	.64	.71	.73	.69	.77	.77
			Delta	11.38	11.52	10.85	10.61	10.99	9.99	10.07
12	11	540.29	Percentage Deviation	16.83	19.62	-23.93	-4.30	7.28	-5.34	-8.22
			Proportion Correct	.80	.82	.59	.64	.71	.69	.75
			Delta	9.67	9.34	12.10	11.55	10.84	10.97	10.35
13	21	374.60	Percentage Deviation	-3.19	0.80	16.10	-14.84	-14.06	16.65	-8.17
			Proportion Correct	.67	.70	.95	.58	.57	.89	.79
			Delta	11.23	10.88	6.53	12.18	12.30	8.20	9.72
14	1	1408.41	Percentage Deviation	-33.09	-32.25	35.67	-25.86	-2.17	40.42	12.45
			Proportion Correct	.37	.38	.99	.40	.50	.90	.90
			Delta	14.33	14.25	3.87	14.00	12.99	7.87	7.81
15	16	440.93	Percentage Deviation	-36.44	-28.93	10.63	5.90	15.66	59.92	-15.63
			Proportion Correct	.21	.23	.45	.34	.37	.57	.38
			Delta	16.29	15.94	13.52	14.66	14.35	12.32	14.26
16	17	432.72	Percentage Deviation	16.37	2.68	-1.11	-21.84	-33.55	45.76	-16.05
			Proportion Correct	.53	.48	.60	.35	.29	.78	.57
			Delta	12.66	13.24	11.98	14.54	15.26	9.97	12.34
17	39	212.26	Percentage Deviation	-3.76	-5.63	16.96	14.09	-7.52	-21.03	-2.62
			Proportion Correct	.49	.49	.69	.58	.46	.44	.60
			Delta	13.05	13.13	11.08	12.23	13.38	13.65	12.00
18	10	542.40	Percentage Deviation	1.98	-3.16	-18.33	-23.26	48.29	24.84	-6.05
			Proportion Correct	.55	.53	.57	.41	.75	.77	.71
			Delta	12.48	12.72	12.35	13.96	10.26	10.01	10.77
19	4	738.52	Percentage Deviation	-32.06	-34.90	22.23	17.07	-8.50	31.94	1.46
			Proportion Correct	.37	.36	.84	.62	.47	.81	.76
			Delta	14.33	14.47	8.97	11.73	13.29	9.46	10.15
20	49	136.61	Percentage Deviation	9.13	12.52	-12.02	-3.19	8.37	0.19	-11.91
			Proportion Correct	.52	.54	.52	.45	.49	.53	.56
			Delta	12.81	12.61	12.84	13.50	13.07	12.68	12.39
21	6	647.54	Percentage Deviation	25.00	22.97	-43.33	-12.14	-24.10	25.64	15.73
			Proportion Correct	.58	.57	.34	.40	.33	.67	.77
			Delta	12.22	12.27	14.64	14.06	14.78	11.25	10.07

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
22	2	1379.09	Percentage Deviation	-37.51	-35.02	27.42	47.82	-26.89	17.76	-6.99
			Proportion Correct	.38	.39	.96	.88	.42	.80	.76
			Delta	14.25	14.08	5.94	8.37	13.84	9.65	10.18
23	3	755.48	Percentage Deviation	-4.95	-13.54	26.75	-39.30	-18.07	66.82	-25.70
			Proportion Correct	.41	.38	.71	.26	.34	.82	.45
			Delta	13.90	14.26	10.84	15.61	14.68	9.36	13.45
24	8	568.74	Percentage Deviation	-5.30	-6.64	42.35	-1.91	13.02	2.07	-40.38
			Proportion Correct	.41	.41	.82	.42	.35	.51	.38
			Delta	13.89	13.92	9.34	13.83	14.49	12.85	14.23
25	5	652.97	Percentage Deviation	-14.01	-18.92	40.90	-5.41	-11.40	-5.16	-7.23
			Proportion Correct	.40	.38	.91	.42	.37	.52	.67
			Delta	14.07	14.25	7.68	13.77	14.28	12.78	11.23
26	7	626.91	Percentage Deviation	22.44	25.19	-6.99	-41.67	-32.00	34.98	-9.35
			Proportion Correct	.59	.61	.59	.27	.30	.75	.63
			Delta	12.10	11.91	12.15	15.43	15.04	10.27	11.72
27	9	556.50	Percentage Deviation	3.77	8.35	-34.58	41.46	-29.11	31.13	-23.52
			Proportion Correct	.46	.49	.35	.62	.30	.64	.43
			Delta	13.36	13.13	14.57	11.79	15.04	11.55	13.66
28	20	375.26	Percentage Deviation	-29.26	-31.15	26.91	-5.10	-21.98	21.52	15.95
			Proportion Correct	.22	.22	.62	.29	.22	.47	.66
			Delta	16.10	16.14	11.79	15.25	16.07	13.28	11.32
29	51	130.24	Percentage Deviation	-8.35	-12.28	8.89	-9.82	-9.78	18.56	5.67
			Proportion Correct	.27	.27	.50	.26	.25	.44	.56
			Delta	15.40	15.49	13.01	15.55	15.71	13.62	12.38
30	31	265.39	Percentage Deviation	-5.06	-5.18	10.39	-12.86	2.87	55.45	-44.10
			Proportion Correct	.13	.13	.21	.11	.14	.25	.12
			Delta	17.50	17.48	16.27	17.82	17.40	15.75	17.72
31	54	105.81	Percentage Deviation	0.06	-2.89	3.76	-1.96	-3.14	-1.31	4.46
			Proportion Correct	.73	.72	.91	.70	.67	.81	.98
			Delta	10.51	10.72	7.58	10.86	11.27	9.54	5.05
32	58	88.24	Percentage Deviation	0.03	-0.90	1.33	-4.77	2.33	2.47	1.10
			Proportion Correct	.89	.89	.96	.84	.90	.95	.98
			Delta	7.99	8.12	5.79	8.95	7.95	6.31	4.65

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
33	18	426.72	Percentage Deviation	19.47	25.96	-11.95	-19.02	-13.56	-22.26	18.86
			Proportion Correct	.62	.66	.57	.41	.43	.45	.84
			Delta	11.76	11.37	12.29	13.87	13.75	13.46	9.04
34	48	136.94	Percentage Deviation	1.38	-0.50	-0.81	3.96	-0.56	-7.47	3.39
			Proportion Correct	.84	.83	.91	.85	.80	.81	.98
			Delta	9.0	9.25	7.77	8.81	9.62	9.50	5.05
35	40	194.98	Percentage Deviation	2.41	1.75	-0.69	5.56	-3.85	-9.31	1.19
			Proportion Correct	.93	.93	.95	.95	.86	.85	.98
			Delta	7.10	7.22	6.46	6.35	8.69	8.85	4.50
36	29	273.01	Percentage Deviation	10.57	9.55	-8.71	-0.35	-4.08	-17.33	9.31
			Proportion Correct	.79	.78	.75	.70	.66	.64	.94
			Delta	9.84	9.89	10.35	10.93	11.41	11.62	6.74
37	46	152.04	Percentage Deviation	4.49	6.47	-2.49	-10.77	4.34	-2.88	2.66
			Proportion Correct	.83	.85	.88	.70	.79	.83	.97
			Delta	9.19	8.91	8.26	10.93	9.72	9.14	5.28
38	52	122.40	Percentage Deviation	5.05	4.26	-6.37	-1.18	5.21	3.27	-8.48
			Proportion Correct	.68	.68	.74	.63	.65	.75	.77
			Delta	11.10	11.12	10.45	11.68	11.50	10.32	10.00
39	34	257.79	Percentage Deviation	-2.84	-0.32	-2.58	16.43	-0.22	-16.16	3.97
			Proportion Correct	.75	.77	.85	.89	.75	.69	.94
			Delta	10.29	10.00	8.90	8.09	10.36	10.98	6.62
40	56	93.11	Percentage Deviation	-1.16	1.67	3.31	-3.99	-3.06	-2.19	5.19
			Proportion Correct	.74	.76	.86	.71	.71	.78	.91
			Delta	10.41	10.21	8.70	10.75	10.80	9.95	7.69
41	53	117.40	Percentage Deviation	-1.70	-2.42	-4.08	4.86	9.36	-5.68	3.71
			Proportion Correct	.70	.70	.82	.73	.73	.74	.94
			Delta	10.95	10.96	9.40	10.56	10.52	10.39	6.74
42	45	172.09	Percentage Deviation	-7.15	-5.18	7.04	2.45	6.80	-7.99	5.40
			Proportion Correct	.70	.72	.92	.76	.78	.75	.95
			Delta	10.90	10.69	7.36	10.13	9.94	10.34	6.43
43	55	102.18	Percentage Deviation	3.45	4.87	-5.07	1.86	-5.53	1.78	-3.65
			Proportion Correct	.86	.87	.85	.84	.77	.89	.89
			Delta	8.66	8.40	8.85	9.02	10.07	8.17	8.18

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
44	59	81.37	Percentage Deviation	5.45	2.97	-4.85	-2.81	-0.79	-1.76	2.98
			Proportion Correct	.67	.66	.77	.60	.59	.72	.91
			Delta	11.26	11.38	10.05	11.95	12.15	10.72	7.73
45	57	92.39	Percentage Deviation	6.86	4.94	-5.76	-2.72	-1.12	0.30	-1.72
			Proportion Correct	.59	.58	.66	.53	.51	.63	.75
			Delta	12.10	12.16	11.33	12.76	12.89	11.67	10.26
46	32	260.72	Percentage Deviation	5.66	5.66	-8.67	-2.17	-2.93	2.33	0.78
			Proportion Correct	.87	.87	.85	.79	.76	.91	.97
			Delta	8.53	8.46	8.91	9.79	10.15	7.58	5.39
47	60	80.49	Percentage Deviation	2.14	-1.15	-1.71	-8.25	6.54	6.86	-0.95
			Proportion Correct	.68	.66	.81	.60	.66	.80	.88
			Delta	11.19	11.39	9.50	12.05	11.35	9.62	8.28
48	43	179.61	Percentage Deviation	-4.52	1.01	10.18	19.29	-12.52	-18.92	-5.19
			Proportion Correct	.48	.51	.70	.59	.42	.46	.66
			Delta	13.21	12.89	10.89	12.10	13.81	13.43	11.34
49	30	271.23	Percentage Deviation	10.75	9.28	-6.95	-12.77	1.52	-1.64	1.04
			Proportion Correct	.35	.84	.82	.66	.75	.82	.93
			Delta	8.94	9.07	9.39	11.42	10.38	9.39	7.06
50	24	344.43	Percentage Deviation	11.78	10.99	-1.59	-5.31	-0.69	-24.98	6.61
			Proportion Correct	.75	.75	.78	.63	.65	.55	.89
			Delta	10.27	10.29	10.00	11.69	11.53	12.51	8.14
51	23	357.21	Percentage Deviation	19.92	19.77	-4.73	-12.83	0.64	-26.96	2.13
			Proportion Correct	.66	.66	.69	.47	.51	.47	.82
			Delta	11.39	11.34	11.01	13.36	12.93	13.34	9.38
52	38	213.97	Percentage Deviation	16.68	14.29	-12.02	-13.31	10.82	-4.75	-2.93
			Proportion Correct	.51	.50	.56	.37	.44	.51	.71
			Delta	12.96	13.02	12.41	14.40	13.70	12.95	10.85
53	27	309.14	Percentage Deviation	15.24	13.14	-1.15	-17.73	6.00	-20.81	4.55
			Proportion Correct	.72	.71	.75	.51	.64	.55	.86
			Delta	10.64	10.75	10.28	12.94	11.62	12.51	8.77
54	36	250.36	Percentage Deviation	10.54	7.99	-2.89	-1.31	-7.05	-11.16	-0.50
			Proportion Correct	.86	.84	.87	.76	.70	.75	.93
			Delta	8.72	8.99	8.58	10.21	10.96	10.31	7.09

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
55	33	258.01	Percentage Deviation	9.32	9.05	-3.54	-4.07	-0.61	-11.58	-0.47
			Proportion Correct	.82	.82	.85	.71	.71	.74	.93
			Delta	9.42	9.41	8.82	10.85	10.83	10.52	6.95
56	47	141.49	Percentage Deviation	3.19	6.29	-3.00	2.95	-1.55	-12.80	3.82
			Proportion Correct	.69	.71	.81	.69	.63	.67	.94
			Delta	11.05	10.76	9.53	11.14	11.77	11.31	6.90
57	41	184.39	Percentage Deviation	-3.58	-9.31	3.63	16.48	-2.00	-23.68	14.15
			Proportion Correct	.41	.39	.64	.50	.40	.39	.80
			Delta	13.94	14.15	11.68	13.13	14.13	14.16	9.76
58	19	412.44	Percentage Deviation	14.13	21.74	-26.92	-2.24	-7.30	-0.11	9.66
			Proportion Correct	.56	.60	.51	.48	.43	.61	.86
			Delta	12.38	11.97	12.94	13.31	13.86	11.98	8.66
59	12	516.41	Percentage Deviation	5.32	6.46	4.97	7.91	0.32	-37.65	7.29
			Proportion Correct	.60	.61	.81	.61	.55	.42	.91
			Delta	12.03	11.91	9.61	11.96	12.69	13.84	7.78
60	28	303.03	Percentage Deviation	8.77	11.06	-11.37	6.52	-4.05	-14.28	2.56
			Proportion Correct	.74	.75	.75	.72	.64	.66	.91
			Delta	10.51	10.29	10.50	10.75	11.77	11.41	7.63

Administration II (Form BTF11)

TOEFL Section I: Listening Comprehension

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
1	45	77.89	Percentage Deviation	0.85	2.27	-3.84	3.57	-1.96	-1.65	0.0
			Proportion Correct	.85	.85	.86	.88	.83	.80	.95
			Delta	8.89	8.80	8.73	8.33	9.23	9.61	6.46
2	46	77.70	Percentage Deviation	1.77	4.73	0.57	-5.69	1.59	-7.13	0.0
			Proportion Correct	.70	.71	.79	.66	.71	.60	.94
			Delta	10.88	10.78	9.77	11.33	10.75	12.00	6.86
3	36	115.38	Percentage Deviation	1.25	1.95	-0.49	6.97	1.41	-27.01	0.0
			Proportion Correct	.68	.68	.76	.73	.69	.47	.89
			Delta	11.13	11.18	10.21	10.55	10.97	13.28	8.12
4	34	122.08	Percentage Deviation	3.16	5.50	5.69	-3.80	-1.73	-22.'2	0.0
			Proportion Correct	.76	.77	.85	.72	.73	.53	.93
			Delta	10.20	10.08	8.81	10.71	10.52	12.45	7.15
5	48	69.11	Percentage Deviation	2.62	3.04	-3.30	-4.20	7.00	-11.04	0.0
			Proportion Correct	.53	.51	.64	.51	.58	.41	.89
			Delta	12.74	12.87	11.60	12.93	12.21	13.96	8.10
6	5	323.57	Percentage Deviation	0.19	-0.94	-2.25	-22.64	11.30	37.54	0.0
			Proportion Correct	.66	.64	.75	.52	.75	.82	.96
			Delta	11.37	11.58	10.25	12.80	10.29	9.28	5.98
7	10	238.99	Percentage Deviation	6.15	12.85	-7.77	11.32	-10.47	-26.86	0.0
			Proportion Correct	.69	.72	.68	.73	.59	.46	.79
			Delta	11.04	10.65	11.18	10.51	12.09	13.44	9.75
8	13	208.50	Percentage Deviation	13.71	9.47	-11.42	-4.32	-12.17	14.71	0.0
			Proportion Correct	.78	.74	.67	.66	.61	.75	.84
			Delta	9.93	10.39	11.20	11.32	11.86	10.27	9.03
9	28	135.78	Percentage Deviation	7.95	8.09	0.18	-4.82	-6.58	-11.26	0.0
			Proportion Correct	.75	.74	.77	.67	.66	.59	.89
			Delta	10.35	10.44	10.00	11.27	11.40	12.11	8.04
10	39	94.41	Percentage Deviation	-5.86	-1.84	3.51	-9.16	12.32	-0.39	0.0
			Proportion Correct	.49	.50	.67	.49	.62	.48	.82
			Delta	13.07	12.96	11.24	13.10	11.80	13.19	9.28

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
11	3	409.23	Percentage Deviation	16.23	16.47	-15.84	4.39	-26.92	22.06	0.0
			Proportion Correct	.70	.69	.59	.64	.45	.69	.93
			Delta	10.87	10.99	12.05	11.53	13.45	11.06	7.12
12	25	140.88	Percentage Deviation	2.93	7.84	-9.69	16.09	-9.02	-14.18	0.0
			Proportion Correct	.51	.52	.57	.59	.47	.39	.85
			Delta	12.92	12.82	12.28	12.07	13.27	14.16	8.84
13	43	82.42	Percentage Deviation	-1.85	-0.34	3.12	13.34	-14.35	-1.74	0.0
			Proportion Correct	.42	.42	.58	.50	.39	.38	.80
			Delta	13.82	13.35	12.16	12.99	14.10	14.19	9.63
14	44	81.24	Percentage Deviation	-1.06	-0.36	-1.27	-0.83	-3.00	17.62	0.0
			Proportion Correct	.64	.63	.77	.66	.65	.68	.95
			Delta	11.53	11.66	10.08	11.32	11.48	11.16	6.54
15	14	200.38	Percentage Deviation	-3.52	-1.43	2.05	-1.11	13.47	-26.69	0.0
			Proportion Correct	.49	.49	.66	.52	.61	.35	.91
			Delta	13.07	13.07	11.32	12.79	11.84	14.57	7.52
16	37	113.78	Percentage Deviation	5.69	5.04	0.83	-13.10	-0.18	5.07	0.0
			Proportion Correct	.54	.52	.66	.45	.54	.49	.91
			Delta	12.63	12.79	11.35	13.45	12.56	13.10	7.68
17	32	123.51	Percentage Deviation	4.14	8.12	4.71	-14.01	-7.49	10.40	0.0
			Proportion Correct	.53	.54	.65	.45	.49	.52	.79
			Delta	12.74	12.65	11.46	13.54	13.15	12.83	9.82
18	33	123.15	Percentage Deviation	1.10	2.23	-2.34	-22.41	21.31	0.32	0.0
			Proportion Correct	.34	.33	.46	.27	.45	.30	.55
			Delta	14.67	14.72	13.37	15.44	13.52	15.06	12.48
19	23	149.26	Percentage Deviation	-8.64	-4.47	-0.36	8.57	-11.41	46.68	0.0
			Proportion Correct	.24	.24	.42	.30	.27	.34	.63
			Delta	15.82	15.76	13.82	15.07	15.49	14.62	11.65
20	9	251.24	Percentage Deviation	-14.69	-14.01	-6.10	15.34	-12.78	93.97	0.0
			Proportion Correct	.25	.25	.41	.36	.29	.50	.72
			Delta	15.64	15.71	13.93	14.42	15.26	13.03	10.70

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
21	31	124.31	Percentage Deviation	-4.74	-5.50	3.39	6.41	1.59	-3.37	0.0
			Proportion Correct	.82	.80	.93	.92	.87	.80	.97
			Delta	9.39	9.57	6.99	7.39	8.45	9.58	5.61
22	40	92.84	Percentage Deviation	1.89	4.47	-2.96	1.82	-5.16	0.01	0.0
			Proportion Correct	.90	.92	.90	.91	.84	.86	.96
			Delta	7.77	7.33	7.96	7.64	9.02	8.60	5.77
23	49	63.37	Percentage Deviation	1.08	-0.89	-0.08	2.78	0.02	-6.98	0.0
			Proportion Correct	.87	.84	.91	.89	.86	.77	.99
			Delta	8.51	8.95	7.63	8.06	8.66	10.07	4.16
24	8	253.88	Percentage Deviation	2.54	3.66	-0.73	8.71	-4.37	-23.71	0.0
			Proportion Correct	.85	.85	.88	.91	.79	.60	.93
			Delta	8.91	8.92	8.24	7.67	9.80	11.95	6.99
25	16	190.77	Percentage Deviation	-6.80	-4.42	1.97	11.04	3.26	-13.60	0.0
			Proportion Correct	.71	.72	.86	.86	.79	.61	.98
			Delta	10.79	10.71	8.61	8.69	9.71	11.81	4.58
26	29	130.99	Percentage Deviation	-0.46	-2.57	1.83	-10.97	10.10	4.42	0.0
			Proportion Correct	.68	.66	.81	.62	.78	.65	.97
			Delta	11.09	11.40	9.45	11.73	9.96	11.41	5.49
27	1	681.05	Percentage Deviation	12.10	12.71	3.89	11.27	-23.05	-42.05	0.0
			Proportion Correct	.84	.84	.87	.85	.59	.42	.97
			Delta	8.96	9.08	8.49	8.85	12.14	13.80	5.72
28	22	153.43	Percentage Deviation	-0.70	0.50	-3.49	1.13	-9.47	30.84	0.0
			Proportion Correct	.68	.68	.76	.71	.63	.83	.96
			Delta	11.12	11.16	10.23	10.82	11.65	9.14	6.21
29	11	237.52	Percentage Deviation	-8.89	-9.43	9.96	-8.44	15.53	-2.35	0.0
			Proportion Correct	.59	.57	.84	.61	.77	.58	.96
			Delta	12.10	12.25	9.01	11.93	9.99	12.16	6.21
30	2	427.81	Percentage Deviation	-12.26	-12.01	11.97	-26.10	20.38	36.67	0.0
			Proportion Correct	.48	.47	.78	.42	.70	.66	.92
			Delta	13.2[illegible]	13.34	9.88	13.84	10.93	11.38	7.44

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
31	24	148.55	Percentage Deviation	-3.52	-3.57	-9.36	4.33	10.08	8.32	0.0
			Proportion Correct	.64	.63	.71	.71	.76	.65	.92
			Delta	11.52	11.68	10.74	10.74	10.22	11.46	7.47
32	7	265.23	Percentage Deviation	-15.10	-28.12	5.62	16.63	17.91	0.67	0.0
			Proportion Correct	.42	.35	.65	.59	.61	.45	.92
			Delta	13.83	14.58	11.50	12.08	11.87	13.48	7.36
33	38	100.57	Percentage Deviation	6.34	9.13	-12.39	-7.96	6.01	2.72	0.0
			Proportion Correct	.58	.58	.57	.51	.59	.52	.67
			Delta	12.19	12.15	12.34	12.86	12.04	12.79	11.19
34	19	175.90	Percentage Deviation	0.07	3.65	11.21	-25.21	7.75	-0.04	0.0
			Proportion Correct	.47	.48	.70	.37	.54	.41	.88
			Delta	13.26	13.24	10.94	14.34	12.56	13.89	8.27
35	20	170.82	Percentage Deviation	-16.54	-24.80	11.82	-6.25	22.74	6.28	0.0
			Proportion Correct	.23	.20	.50	.28	.40	.26	.75
			Delta	15.91	16.33	12.97	15.35	14.06	15.62	10.28
36	4	350.17	Percentage Deviation	-25.82	-23.38	8.84	29.45	8.83	-3.01	0.0
			Proportion Correct	.40	.41	.71	.71	.61	.49	.84
			Delta	14.01	13.95	10.84	10.74	11.89	13.10	8.94
37	30	129.10	Percentage Deviation	-3.39	-7.81	2.62	4.93	6.62	-8.50	0.0
			Proportion Correct	.75	.70	.88	.83	.83	.66	.97
			Delta	10.33	10.87	8.24	9.26	9.11	11.31	5.16
38	21	159.56	Percentage Deviation	-5.04	-5.58	-8.47	17.36	6.15	-9.25	0.0
			Proportion Correct	.53	.51	.61	.67	.61	.46	.73
			Delta	12.72	12.87	11.93	11.26	11.87	13.39	10.59
39	15	199.74	Percentage Deviation	5.80	7.22	1.96	-8.99	8.25	-36.88	0.0
			Proportion Correct	.65	.64	.75	.57	.69	.36	.88
			Delta	11.47	11.53	10.32	12.27	11.02	14.42	8.30
40	35	118.62	Percentage Deviation	-2.10	-7.25	0.15	-3.22	1.75	26.24	0.0
			Proportion Correct	.64	.59	.76	.64	.68	.75	.93
			Delta	11.61	12.07	10.14	11.54	11.14	10.27	7.15

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
41	47	73.79	Percentage Deviation	2.46	-3.17	2.94	-5.26	-0.11	6.65	0.0
			Proportion Correct	.66	.61	.78	.63	.67	.64	.93
			Delta	11.33	11.84	9.87	11.72	11.26	11.58	7.15
42	12	226.16	Percentage Deviation	6.42	1.59	6.18	-14.59	-9.72	23.96	0.0
			Proportion Correct	.71	.67	.82	.59	.62	.77	.93
			Delta	10.73	11.23	9.30	12.14	11.78	10.10	6.96
43	6	269.75	Percentage Deviation	8.53	5.42	-10.59	14.92	-17.49	1.98	0.0
			Proportion Correct	.70	.67	.66	.75	.54	.61	.86
			Delta	10.92	11.26	11.30	10.25	12.57	11.84	8.71
44	17	189.65	Percentage Deviation	-9.07	-4.40	-13.05	41.73	-2.84	-27.18	0.0
			Proportion Correct	.32	.33	.37	.51	.35	.25	.62
			Delta	14.88	14.75	14.29	12.90	14.49	15.73	11.80
45	41	84.16	Percentage Deviation	-1.05	-0.30	3.21	-5.80	-6.13	23.63	0.0
			Proportion Correct	.47	.47	.59	.46	.46	.55	.70
			Delta	13.29	13.32	12.13	13.42	13.36	12.47	10.92
46	26	140.41	Percentage Deviation	16.25	9.93	10.86	-5.43	-20.11	-29.02	0.0
			Proportion Correct	.48	.45	.49	.40	.33	.29	.57
			Delta	13.17	13.46	13.10	14.05	14.73	15.22	12.31
47	18	179.93	Percentage Deviation	-8.39	-11.01	3.99	-5.77	4.59	38.89	0.0
			Proportion Correct	.52	.49	.73	.55	.62	.71	.94
			Delta	12.84	13.10	10.58	12.55	11.80	10.83	6.90
48	27	136.29	Percentage Deviation	-3.74	-7.36	18.68	-2.69	-1.33	-17.90	0.0
			Proportion Correct	.41	.39	.63	.43	.45	.34	.68
			Delta	13.87	14.09	11.72	13.72	13.54	14.69	11.11
49	50	56.19	Percentage Deviation	-2.06	-1.62	-4.72	3.47	3.42	5.59	0.0
			Proportion Correct	.46	.46	.60	.51	.53	.46	.82
			Delta	13.36	13.45	12.03	12.92	12.75	13.41	9.27
50	42	83.12	Percentage Deviation	1.21	6.76	1.61	-11.14	-3.00	13.18	0.0
			Proportion Correct	.4	.49	.62	.43	.49	.49	.82
			Delta	13.2	13.10	11.83	13.71	13.17	13.17	9.37

Administration II (Form BTF11)

TOEFL Section II: Structure and Written Expression

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
1	17	203.50	Percentage Deviation	-7.78	-8.15	-0.33	7.87	6.04	3.75	0.73
			Proportion Correct	.78	.78	.86	.92	.86	.94	.96
			Delta	9.86	9.93	8.63	7.41	8.66	6.78	5.93
2	29	127.84	Percentage Deviation	4.56	3.47	-2.16	-2.90	-0.42	-2.55	1.27
			Proportion Correct	.92	.91	.87	.86	.84	.90	.95
			Delta	7.28	7.62	8.49	8.72	9.02	7.83	6.58
3	31	121.23	Percentage Deviation	-4.01	-6.26	3.46	2.58	1.92	2.20	1.22
			Proportion Correct	.84	.82	.92	.90	.86	.95	.98
			Delta	8.99	9.37	7.42	7.86	8.75	6.55	5.09
4	14	277.94	Percentage Deviation	6.41	2.57	9.41	9.82	-12.41	-16.06	-7.78
			Proportion Correct	.79	.75	.84	.82	.60	.69	.82
			Delta	9.82	10.25	9.08	9.41	11.95	11.06	9.33
5	37	84.58	Percentage Deviation	2.22	3.11	-0.83	1.93	-4.82	-3.74	-0.11
			Proportion Correct	.93	.93	.90	.92	.82	.91	.97
			Delta	7.20	7.04	7.85	7.31	9.28	7.67	5.36
6	27	140.70	Percentage Deviation	0.25	-0.30	-8.69	9.93	-5.62	2.87	2.30
			Proportion Correct	.74	.73	.70	.82	.64	.85	.93
			Delta	10.40	10.53	10.91	9.40	11.54	8.50	7.09
7	21	174.79	Percentage Deviation	-0.75	-0.52	11.71	-11.41	1.73	-1.27	-0.36
			Proportion Correct	.75	.75	.88	.68	.74	.82	.90
			Delta	10.27	10.29	8.26	11.18	10.46	9.35	7.83
8	34	115.43	Percentage Deviation	5.95	5.30	-2.07	-7.47	5.89	-12.10	-0.28
			Proportion Correct	.64	.63	.63	.56	.58	.62	.81
			Delta	11.62	11.72	11.64	12.39	12.15	11.79	9.42
9	24	155.93	Percentage Deviation	-4.64	-7.70	-7.11	11.45	1.04	10.30	1.40
			Proportion Correct	.65	.62	.67	.77	.62	.89	.92
			Delta	11.43	11.73	11.24	10.06	11.80	8.17	7.44
10	35	97.84	Percentage Deviation	-3.23	-2.97	-2.58	8.19	-5.96	1.90	3.23
			Proportion Correct	.57	.57	.63	.65	.48	.76	.91
			Delta	12.25	12.34	11.65	11.44	13.18	10.13	7.52

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
11	6	460.34	Percentage Deviation	-23.98	-23.09	18.11	-6.71	16.43	12.87	8.03
			Proportion Correct	.43	.43	.75	.54	.59	.81	.93
			Delta	13.66	13.68	10.33	12.58	12.06	9.48	7.12
12	4	521.17	Percentage Deviation	-8.46	-6.88	29.10	-11.84	-21.01	-8.85	11.15
			Proportion Correct	.49	.48	.77	.48	.35	.64	.96
			Delta	13.15	13.16	10.02	13.23	14.49	11.55	6.07
13	23	157.98	Percentage Deviation	-3.31	3.63	-8.56	0.57	5.04	-10.58	6.70
			Proportion Correct	.39	.41	.43	.42	.39	.47	.72
			Delta	14.09	13.86	13.66	13.84	14.09	13.28	10.67
14	9	384.00	Percentage Deviation	26.36	25.18	-8.11	-12.53	-24.74	-18.40	1.55
			Proportion Correct	.61	.59	.49	.42	.31	.50	.76
			Delta	11.92	12.06	13.07	13.76	14.95	13.03	10.17
15	1	629.26	Percentage Deviation	-41.35	-42.80	37.05	-13.64	13.71	13.67	15.47
			Proportion Correct	.21	.20	.62	.32	.37	.58	.81
			Delta	16.25	16.35	11.80	14.89	14.31	12.23	9.45
16	8	416.81	Percentage Deviation	6.09	5.61	-5.73	7.07	-5.85	-23.80	3.59
			Proportion Correct	.90	.90	.81	.91	.77	.68	.95
			Delta	7.82	7.97	9.55	7.64	10.12	11.16	6.34
17	18	193.83	Percentage Deviation	-6.83	-8.45	2.53	1.53	7.60	3.97	1.89
			Proportion Correct	.79	.78	.89	.87	.87	.95	.98
			Delta	9.73	9.96	8.12	8.56	8.50	6.47	4.77
18	15	251.99	Percentage Deviation	3.30	2.47	-10.73	4.10	3.37	1.16	-2.78
			Proportion Correct	.94	.93	.81	.94	.90	.96	.95
			Delta	6.89	7.22	9.48	6.72	7.99	6.20	6.54
19	13	289.41	Percentage Deviation	5.23	6.25	-8.77	4.66	0.73	-9.58	-3.28
			Proportion Correct	.92	.93	.80	.92	.83	.84	.93
			Delta	7.31	7.16	9.60	7.49	9.15	9.10	7.15
20	20	180.17	Percentage Deviation	4.45	1.17	-10.62	2.23	1.61	2.15	0.26
			Proportion Correct	.91	.88	.79	.89	.83	.96	.98
			Delta	7.60	8.35	9.85	8.06	9.28	6.20	4.77

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
21	40	64.38	Percentage Deviation	-2.69	-2.33	0.91	-0.17	1.37	3.93	1.00
			Proportion Correct	.84	.83	.88	.86	.81	.96	.98
			Delta	9.09	9.11	8.39	8.72	9.45	6.10	5.01
22	39	67.67	Percentage Deviation	0.55	-1.15	-1.64	-1.78	3.52	0.61	0.45
			Proportion Correct	.79	.77	.79	.78	.76	.88	.94
			Delta	9.75	10.02	9.73	9.98	10.28	8.39	6.69
23	36	95.42	Percentage Deviation	1.38	1.25	-5.40	0.13	-2.11	3.37	2.54
			Proportion Correct	.77	.77	.74	.77	.68	.88	.96
			Delta	9.99	10.08	10.39	10.08	11.13	8.21	6.21
24	28	131.04	Percentage Deviation	3.65	2.50	-0.26	1.16	-4.79	-4.65	-0.40
			Proportion Correct	.88	.87	.85	.86	.74	.88	.97
			Delta	8.26	8.56	8.81	8.69	10.47	8.30	5.66
25	7	454.82	Percentage Deviation	8.90	9.88	-9.70	13.62	-6.80	-20.25	-5.72
			Proportion Correct	.86	.86	.73	.90	.69	.68	.86
			Delta	8.68	8.62	10.60	7.93	11.03	11.14	8.76
26	2	554.69	Percentage Deviation	10.35	15.38	-10.96	3.71	1.80	4.26	-20.08
			Proportion Correct	.82	.86	.67	.77	.71	.84	.68
			Delta	9.30	8.76	11.23	10.02	10.85	9.10	11.16
27	11	352.45	Percentage Deviation	21.49	20.48	-15.66	-7.01	-13.49	-2.11	-4.55
			Proportion Correct	.80	.79	.57	.61	.54	.70	.75
			Delta	9.64	9.79	12.29	11.87	12.70	10.89	10.33
28	3	549.99	Percentage Deviation	-1.58	1.71	-8.01	19.45	1.00	-28.67	2.01
			Proportion Correct	.72	.74	.70	.88	.67	.60	.94
			Delta	10.63	10.42	10.95	8.28	11.35	12.02	6.81
29	26	142.89	Percentage Deviation	5.89	6.12	3.59	-2.03	-15.47	0.72	0.03
			Proportion Correct	.75	.74	.77	.70	.55	.81	.90
			Delta	10.35	10.42	10.10	10.96	12.55	9.48	7.98
30	25	146.73	Percentage Deviation	-9.37	-9.29	11.34	-2.55	-0.01	6.14	3.77
			Proportion Correct	.52	.51	.71	.57	.53	.74	.86
			Delta	12.82	12.87	10.87	12.34	12.81	10.41	8.63

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
31	10	358.72	Percentage Deviation	13.80	16.06	-31.10	-6.09	12.96	7.40	-5.55
			Proportion Correct	.67	.68	.43	.55	.62	.71	.70
			Delta	11.26	11.17	13.78	12.47	11.82	10.78	10.91
32	38	69.76	Percentage Deviation	0.73	1.54	-1.62	-5.74	2.04	6.66	-0.07
			Proportion Correct	.67	.67	.70	.63	.63	.83	.89
			Delta	11.26	11.27	10.91	11.64	11.76	9.14	8.11
33	19	183.59	Percentage Deviation	-9.14	-8.08	10.30	-0.23	-5.91	20.23	-1.28
			Proportion Correct	.42	.42	.61	.48	.39	.76	.80
			Delta	13.79	13.82	11.97	13.23	14.23	10.13	9.65
34	33	116.91	Percentage Deviation	-8.46	-7.17	2.61	0.34	7.66	5.95	1.43
			Proportion Correct	.47	.47	.60	.53	[illegible]	.70	.82
			Delta	13.27	13.27	12.05	12.72	13.11	10.89	9.38
35	30	122.55	Percentage Deviation	-0.39	-4.21	-10.12	14.35	-0.47	13.43	-3.69
			Proportion Correct	.45	.43	.48	.53	.42	.66	.71
			Delta	13.51	13.73	13.33	12.74	13.94	11.39	10.87
36	12	308.06	Percentage Deviation	19.57	14.69	-0.50	-28.15	-3.30	-16.44	5.17
			Proportion Correct	.61	.58	.60	.38	.46	.56	.88
			Delta	11.89	12.24	12.15	14.27	13.62	12.37	8.32
37	16	223.70	Percentage Deviation	-12.14	-3.32	30.04	-8.26	-19.13	12.45	-1.78
			Proportion Correct	.32	.34	.59	.34	.28	.56	.66
			Delta	14.90	14.60	12.26	14.64	15.58	12.42	11.48
38	22	161.12	Percentage Deviation	3.59	-1.97	-8.72	-1.95	11.03	36.47	-12.33
			Proportion Correct	.36	.33	.40	.35	.38	.65	.55
			Delta	14.48	14.74	14.17	14.55	14.46	11.51	12.56
39	32	121.14	Percentage Deviation	-9.60	-0.28	7.55	-13.75	14.00	-17.05	6.72
			Proportion Correct	.29	.31	.44	.28	.35	.36	.64
			Delta	15.27	14.97	13.75	15.33	14.72	14.41	11.59
40	5	492.24	Percentage Deviation	-47.68	-39.65	22.72	-32.37	-3.61	97.99	10.87
			Proportion Correct	.10	.11	.34	.13	.20	.56	.49
			Delta	18.20	17.91	14.76	17.52	16.62	12.38	13.14

Administration II (Form BTF11)

TOEFL Section III: Reading Comprehension and Vocabulary

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
1	59	69.63	Percentage Deviation	-0.90	0.57	-0.44	-1.53	3.25	-2.98	0.57
			Proportion Correct	.91	.92	.92	.87	.89	.90	.97
			Delta	7.68	7.36	7.26	8.54	8.17	7.83	5.72
2	20	272.43	Percentage Deviation	2.76	3.50	2.63	-9.33	-2.06	1.30	1.11
			Proportion Correct	.96	.97	.97	.80	.83	.97	.99
			Delta	5.81	5.54	5.30	9.65	9.20	5.54	2.75
3	11	442.39	Percentage Deviation	-7.27	-7.30	9.89	-5.68	7.98	-8.65	6.33
			Proportion Correct	.78	.78	.95	.73	.79	.79	.99
			Delta	9.92	9.95	6.36	10.54	9.73	9.77	4.27
4	36	178.06	Percentage Deviation	1.84	-1.68	4.24	-12.79	11.11	1.67	-2.89
			Proportion Correct	.70	.68	.76	.55	.66	.73	.78
			Delta	10.85	11.16	10.15	12.49	11.30	10.54	9.85
5	23	251.47	Percentage Deviation	-2.54	-2.37	9.74	-9.33	-2.81	5.46	3.27
			Proportion Correct	.85	.85	.98	.72	.74	.94	.98
			Delta	8.92	8.92	5.03	10.61	10.48	6.71	4.77
6	6	591.17	Percentage Deviation	2.93	5.66	-21.27	5.78	1.40	8.91	2.51
			Proportion Correct	.88	.90	.69	.82	.73	.96	.97
			Delta	8.32	7.86	10.99	9.37	10.51	5.89	5.43
7	46	134.51	Percentage Deviation	-3.74	-3.83	3.41	-1.67	3.76	4.78	0.08
			Proportion Correct	.88	.87	.95	.85	.87	.97	.96
			Delta	8.38	8.43	6.23	8.88	8.58	5.41	5.93
8	19	306.45	Percentage Deviation	-1.08	1.04	3.00	11.13	-14.84	30.92	-13.54
			Proportion Correct	.66	.67	.74	.64	.45	.92	.71
			Delta	11.35	11.20	10.45	11.54	13.53	7.26	10.82
9	38	176.81	Percentage Deviation	-7.55	-9.98	9.12	-2.66	14.46	3.71	-3.20
			Proportion Correct	.63	.61	.79	.60	.66	.73	.78
			Delta	11.69	11.87	9.80	12.00	11.35	10.49	9.88
10	4	687.51	Percentage Deviation	4.12	3.15	11.98	-1.30	-0.59	15.10	-22.95
			Proportion Correct	.81	.80	.91	.69	.65	.94	.68
			Delta	9.45	9.57	7.65	11.02	11.51	6.93	11.11

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
11	10	480.67	Percentage Deviation	12.23	9.10	-25.28	16.53	-6.35	21.50	-13.28
			Proportion Correct	.78	.75	.54	.74	.57	.87	.68
			Delta	9.97	10.28	12.59	10.39	12.31	8.52	11.08
12	2	773.45	Percentage Deviation	12.78	11.45	19.67	0.92	-11.75	15.33	-39.96
			Proportion Correct	.75	.73	.81	.61	.50	.79	.44
			Delta	10.36	10.49	9.47	11.90	13.01	9.83	13.65
13	18	314.35	Percentage Deviation	4.26	6.25	7.04	-6.29	5.39	-39.84	1.95
			Proportion Correct	.74	.76	.79	.60	.63	.45	.82
			Delta	10.38	10.23	9.79	11.98	11.70	13.55	9.41
14	53	97.29	Percentage Deviation	-0.63	-2.28	1.81	2.06	3.36	-1.38	-2.61
			Proportion Correct	.78	.76	.83	.71	.67	.80	.88
			Delta	9.95	10.14	9.14	10.77	11.28	9.58	8.32
15	29	230.72	Percentage Deviation	-6.37	-3.82	2.06	21.02	-13.44	8.10	-3.16
			Proportion Correct	.70	.72	.81	.79	.51	.85	.86
			Delta	10.89	10.70	9.54	9.80	12.88	8.84	8.67
16	9	487.21	Percentage Deviation	-3.51	-1.78	1.27	-37.15	21.25	49.00	-3.40
			Proportion Correct	.57	.58	.65	.31	.54	.94	.73
			Delta	12.30	12.22	11.41	14.96	12.64	6.78	10.51
17	16	354.39	Percentage Deviation	-6.49	-10.35	4.21	-26.78	17.13	26.54	7.84
			Proportion Correct	.62	.59	.74	.41	.59	.89	.89
			Delta	11.80	12.10	10.40	13.89	12.08	8.17	8.17
18	27	235.65	Percentage Deviation	2.40	6.28	5.40	-12.49	-22.36	32.08	-2.24
			Proportion Correct	.62	.64	.70	.45	.35	.86	.76
			Delta	11.77	11.55	10.92	13.52	14.50	8.76	10.16
19	14	364.97	Percentage Deviation	-21.80	-17.23	-1.16	27.67	19.42	-50.95	18.23
			Proportion Correct	.34	.35	.48	.49	.43	.22	.69
			Delta	14.68	14.49	13.16	13.11	13.70	16.02	10.98
20	25	245.79	Percentage Deviation	-5.43	-3.21	3.44	21.94	10.74	-29.35	-6.67
			Proportion Correct	.55	.55	.65	.60	.50	.43	.68
			Delta	12.54	12.46	11.47	11.95	13.05	13.68	11.09

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
21	1	1039.50	Percentage Deviation	-17.31	-19.18	63.45	-21.77	5.51	42.61	-31.57
			Proportion Correct	.41	.40	.89	.34	.41	.75	.44
			Delta	13.91	14.02	8.07	14.70	13.87	10.27	13.65
22	3	735.78	Percentage Deviation	-24.02	-24.25	27.32	-0.78	-17.34	31.29	13.96
			Proportion Correct	.47	.48	.85	.52	.39	.86	.89
			Delta	13.32	13.35	8.77	12.82	14.16	8.64	8.01
23	8	490.77	Percentage Deviation	6.49	10.28	-46.85	19.00	-23.36	27.14	18.91
			Proportion Correct	.50	.52	.28	.49	.29	.64	.72
			Delta	12.97	12.84	15.39	13.13	15.25	11.60	10.66
24	7	544.43	Percentage Deviation	32.19	26.07	-34.32	7.97	-13.30	22.41	-22.60
			Proportion Correct	.63	.60	.34	.44	.31	.62	.48
			Delta	11.67	11.99	14.61	13.65	14.96	11.74	13.23
25	12	387.32	Percentage Deviation	-7.14	-5.28	33.70	25.49	-20.92	38.27	-39.36
			Proportion Correct	.33	.33	.52	.40	.24	.52	.27
			Delta	14.78	14.73	12.76	14.00	15.81	12.83	15.41
26	17	344.92	Percentage Deviation	16.03	15.85	36.78	-21.71	7.80	-52.31	-27.72
			Proportion Correct	.40	.40	.53	.24	.32	.18	.33
			Delta	13.96	14.00	12.71	15.77	14.90	16.72	14.80
27	33	185.35	Percentage Deviation	11.07	19.09	4.26	9.11	-10.46	-51.68	-8.04
			Proportion Correct	.39	.42	.40	.34	.26	.18	.40
			Delta	14.12	13.85	14.04	14.63	15.55	16.69	13.99
28	15	359.86	Percentage Deviation	-1.54	1.21	-43.38	15.32	2.67	-4.84	27.00
			Proportion Correct	.33	.33	.23	.31	.24	.35	.65
			Delta	14.77	14.71	16.01	14.97	15.78	14.54	11.43
29	13	370.48	Percentage Deviation	0.77	-13.72	-6.46	-33.78	19.91	112.26	-13.88
			Proportion Correct	.19	.16	.23	.11	.18	.46	.29
			Delta	16.45	16.92	15.90	18.01	16.72	13.39	15.25
30	42	167.51	Percentage Deviation	-21.95	-18.11	39.37	5.52	-14.01	-9.13	4.17
			Proportion Correct	.20	.21	.41	.24	.19	.25	.37
			Delta	16.37	16.26	13.87	15.81	16.56	15.70	14.29

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
31	60	69.20	Percentage Deviation	0.31	0.53	-0.29	0.38	0.96	-3.06	-0.43
			Proportion Correct	.95	.95	.95	.93	.92	.93	.97
			Delta	6.31	6.27	6.32	7.11	7.41	7.13	5.49
32	50	108.96	Percentage Deviation	-1.24	-2.48	-2.19	5.71	-1.30	-1.32	2.26
			Proportion Correct	.90	.88	.90	.91	.82	.91	.98
			Delta	7.93	8.25	7.82	7.65	9.36	7.56	4.38
33	49	116.12	Percentage Deviation	-5.02	-2.07	0.06	4.19	-1.21	1.86	3.17
			Proportion Correct	.84	.86	.90	.87	.79	.92	.98
			Delta	9.07	8.65	7.82	8.59	9.79	7.44	4.68
34	5	614.62	Percentage Deviation	4.08	3.85	-27.34	7.78	5.27	10.32	3.30
			Proportion Correct	.85	.85	.61	.82	.77	.93	.93
			Delta	8.80	8.86	11.85	9.33	10.12	7.13	7.06
35	56	80.70	Percentage Deviation	0.18	1.26	-1.00	2.81	-2.94	-3.91	1.36
			Proportion Correct	.90	.91	.91	.88	.80	.88	.97
			Delta	7.77	7.60	7.70	8.36	9.67	8.21	5.55
36	21	259.22	Percentage Deviation	13.68	11.59	-7.48	-2.34	1.89	-17.39	-7.84
			Proportion Correct	.79	.77	.67	.61	.60	.59	.74
			Delta	9.81	10.02	11.22	11.86	12.00	12.04	10.46
37	58	72.01	Percentage Deviation	-0.05	0.02	-0.05	2.42	-3.21	-1.27	1.19
			Proportion Correct	.91	.90	.92	.86	.77	.92	.98
			Delta	7.72	7.77	7.34	8.66	10.00	7.44	4.68
38	52	101.09	Percentage Deviation	0.88	1.56	0.31	3.68	-7.10	-4.87	2.07
			Proportion Correct	.88	.88	.89	.82	.69	.85	.97
			Delta	8.40	8.33	8.03	9.31	11.00	8.84	5.23
39	47	124.49	Percentage Deviation	5.95	3.36	-7.48	1.16	5.34	-12.18	-1.37
			Proportion Correct	.77	.75	.70	.67	.66	.66	.81
			Delta	10.01	10.29	10.90	11.20	11.33	11.34	9.47
40	34	184.75	Percentage Deviation	0.08	-2.13	4.83	8.86	-13.25	-6.02	2.60
			Proportion Correct	.83	.81	.90	.80	.60	.81	.96
			Delta	9.22	9.51	7.91	9.57	12.08	9.51	6.12

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
41	51	101.77	Percentage Deviation	-1.36	-0.88	2.18	2.36	-2.77	-4.03	1.90
			Proportion Correct	.82	.82	.88	.76	.68	.83	.95
			Delta	9.33	9.30	8.34	10.13	11.22	9.21	6.42
42	37	177.00	Percentage Deviation	-6.05	-5.16	-2.35	-10.34	5.29	13.55	10.09
			Proportion Correct	.64	.64	.72	.52	.55	.82	.93
			Delta	11.55	11.52	10.72	12.78	12.52	9.31	7.12
43	22	258.30	Percentage Deviation	-9.96	-11.67	-12.44	17.67	13.31	10.96	4.88
			Proportion Correct	.55	.54	.59	.59	.49	.73	.85
			Delta	12.50	12.63	12.08	12.14	13.16	10.54	8.84
44	39	175.54	Percentage Deviation	3.23	6.98	-11.57	-5.95	-4.97	-26.84	16.82
			Proportion Correct	.39	.40	.40	.28	.25	.30	.68
			Delta	14.16	14.06	14.07	15.30	15.72	15.09	11.14
45	55	87.17	Percentage Deviation	4.58	6.47	1.87	0.32	-14.03	-7.02	0.78
			Proportion Correct	.59	.60	.63	.49	.38	.56	.72
			Delta	12.09	11.99	11.71	13.12	14.23	12.43	10.65
46	45	138.21	Percentage Deviation	-18.05	-6.94	15.00	-10.82	4.99	-19.18	15.45
			Proportion Correct	.25	.28	.41	.23	.26	.26	.51
			Delta	15.75	15.36	13.93	15.96	15.69	15.54	12.90
47	31	203.78	Percentage Deviation	4.26	5.15	3.49	9.05	-24.46	27.47	6.74
			Proportion Correct	.56	.56	.63	.46	.27	.42	.81
			Delta	12.42	12.40	11.72	13.40	15.51	13.78	9.56
48	30	208.64	Percentage Deviation	-18.62	-24.45	0.79	-1.59	9.36	23.11	16.48
			Proportion Correct	.24	.23	.37	.25	.26	.41	.54
			Delta	15.76	16.02	14.39	15.73	15.69	13.96	12.58
49	54	90.66	Percentage Deviation	1.94	3.77	-1.51	-6.17	-2.36	-4.09	4.00
			Proportion Correct	.73	.74	.76	.57	.53	.73	.92
			Delta	10.60	10.48	10.26	12.38	12.88	10.59	7.43
50	24	249.17	Percentage Deviation	-1.57	5.29	-12.10	11.69	-7.42	-36.28	16.94
			Proportion Correct	.43	.46	.45	.42	.32	.30	.75
			Delta	13.67	13.43	13.54	13.89	15.06	15.08	10.33

Item No.	Chi-Square Rank-Order	Chi-Square		Chinese 1	Chinese 2	Spanish	Japanese	Arabic	African	Germanic
51	48	120.90	Percentage Deviation	1.09	3.04	-11.91	4.30	-2.82	5.18	3.85
			Proportion Correct	.67	.68	.63	.59	.49	.74	.87
			Delta	11.29	11.18	11.70	12.18	13.24	10.44	8.51
52	28	234.40	Percentage Deviation	8.27	5.39	-2.92	7.94	-21.67	-21.31	6.68
			Proportion Correct	.67	.65	.67	.56	.35	.53	.89
			Delta	11.25	11.49	11.28	12.50	14.65	12.76	8.12
53	40	172.66	Percentage Deviation	4.28	3.55	-12.77	-17.26	8.61	-2.18	11.04
			Proportion Correct	.46	.45	.46	.30	.34	.48	.75
			Delta	13.40	13.48	13.50	15.21	14.85	13.25	10.38
54	41	170.41	Percentage Deviation	8.23	7.87	-6.49	-4.28	-6.57	-12.63	4.60
			Proportion Correct	.73	.72	.69	.56	.50	.63	.90
			Delta	10.57	10.65	11.08	12.46	13.18	11.72	7.99
55	43	159.63	Percentage Deviation	18.89	9.93	15.30	-27.17	-0.29	-9.22	-13.30
			Proportion Correct	.37	.34	.42	.20	.25	.31	.39
			Delta	14.36	14.68	13.90	16.49	15.86	15.07	14.17
56	57	72.26	Percentage Deviation	3.82	9.59	-0.77	-10.97	7.88	-9.30	-3.77
			Proportion Correct	.45	.47	.49	.35	.40	.42	.55
			Delta	13.52	13.30	13.25	14.64	14.21	13.83	12.57
57	32	195.50	Percentage Deviation	-10.19	-4.94	11.61	-26.40	15.35	-7.62	11.74
			Proportion Correct	.43	.46	.64	.30	.42	.49	.79
			Delta	13.68	13.46	11.74	15.21	14.05	13.12	9.96
58	44	146.20	Percentage Deviation	5.91	5.18	-6.42	4.09	-8.35	-9.26	2.95
			Proportion Correct	.78	.78	.75	.72	.61	.71	.89
			Delta	9.88	9.97	10.44	10.74	12.09	10.85	8.14
59	26	241.32	Percentage Deviation	8.72	9.15	-1.82	-14.75	0.38	-15.23	3.39
			Proportion Correct	.81	.82	.82	.58	.64	.68	.94
			Delta	9.47	9.42	9.55	12.27	11.77	11.16	6.77
60	35	183.18	Percentage Deviation	-3.35	2.67	6.93	0.14	-7.24	-23.59	8.52
			Proportion Correct	.64	.68	.80	.61	.53	.54	.90
			Delta	11.62	11.20	9.87	12.04	12.95	12.62	8.04

TOEFL Research Reports

The Performance of Native Speakers of English on the Test of English as a Foreign Language: Clark, John L.D. Report 1. November 1977.

Discusses the results of the administration of TOEFL to native speakers of English just prior to their graduation from a college-preparatory high school program. Total test score distributions were highly negatively skewed, reinforcing findings of earlier studies that TOEFL is not psychometrically appropriate for discriminating among native speakers of English with respect to English language competence.

An Evaluation of Alternative Item Formats for Testing English as a Foreign Language: Pike, Lewis W. Report 2. June 1979.

Describes an extensive research study conducted from 1972 to 1974 that was designed to explore possible changes in the format and content of TOEFL. Questions of validation, criterion selection, and content specifications were investigated. The report includes the results of these findings and discusses the implications for TOEFL content specifications and internal structure. This study contributed to the restructuring of TOEFL beginning in 1976.

The Performance of Non-Native Speakers of English on TOEFL and Verbal Aptitude Tests: Angelis, Paul J.; Swinton, Spencer S.; and Cowell, William R. Report 3. October 1979.

Gives the results of a study in which 400 graduate and undergraduate applicants took TOEFL and either the GRE Verbal or the SAT Verbal and the Test of Standard Written English (TSWE). Included in the report are comparative data on performance across tests and interpretive information on how combined test results might best be used in the admission process.

An Exploration of Speaking Proficiency Measures in the TOEFL Context: Clark, John L.D., and Swinton, Spencer S. Report 4. October 1979.

Describes a three-year study involving the development and experimental administration of test formats and item types aimed at measuring the English-speaking proficiency of nonnative speakers. Factor analysis and other techniques were used to identify subsets of item formats and individual items having satisfactory correlations with the Foreign Service Institute criterion interview administered to the test subjects. The results were grouped into a prototype "Test of Spoken English."

The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language: Powers, Donald E. Report 5. December 1980.

Summarizes analyses indicating performance of 6,000 nonnative speakers of English on TOEFL and GMAT. In addition to comparisons between native and nonnative speakers, data are included showing performance by language background. A variety of analyses support the basic differences in the two tests by showing expected GMAT verbal scores for various levels of TOEFL scores.

Factor Analysis of the Test of English as a Foreign Language for Several Language Groups: Powers, Donald E., and Swinton, Spencer S. Report 6. December 1980.

Provides evidence from a set of exploratory analytical techniques that three major factors underlie performance on TOEFL. Some support is also found for concluding that these factors may be interpreted differently for several language groups. The report discusses implications for making inferences based on TOEFL subscores and considerations for future test development.

The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings: Clark, John L.D., and Swinton, Spencer S. Report 7. December 1980.

Presents the results of a study that examined the performance of foreign teaching assistants on the Test of Spoken English in relation to their classroom performance as judged by students. Also includes, for purposes of comparison, data showing performance of the same groups of teaching assistants on the Foreign Service oral interview and on TOEFL. Based on the analyses conducted in the study, TSE is shown to be a valid predictor of language abilities for nonnative English-speaking graduate teaching assistants.

Effects of Item Disclosure on TOEFL Performance: Hale, Gordon A.; Angelis, Paul J.; and Thibodeau, Lawrence A. Report 8. December 1980.

Reports the findings of a study designed to examine the effects of performance on TOEFL when a subset of items have been disclosed prior to an administration. Based on data from 16 intensive English training programs, the results indicate significant increases in performance in proportion to the number of items made available to students. Details are provided showing separate results by language group and by item type.

Item Performance Across Native Language Groups on the Test of English as a Foreign Language: Alderman, Donald L., and Holland, Paul W. Report 9. August 1981.

Examines the performance of different native language groups on TOEFL items. Based on the item responses of seven language groups on two test forms, the analyses for test examinees with comparable scores show significant differences in their item performance according to their respective native languages. Specialists in English as a second language analyzed these differences from the point of view of similarities and dissimilarities between the examinees' native language and the English language. Discusses implications for the interpretation and examination of item performance by groups.

For further information about any of the TOEFL Research Reports, write to TOEFL Program Office, Box 899, Princeton, NJ 08541, USA.